

IN-32  
(1+33)

CR

P-3(3)

# The Telecommunications and Data Acquisition Progress Report 42-87

## July – September 1986

E. C. Posner  
Editor

November 15, 1986



National Aeronautics and  
Space Administration

Jet Propulsion Laboratory  
California Institute of Technology  
Pasadena, California

(NASA-CR-179938) THE TELECOMMUNICATIONS AND  
DATA ACQUISITION REPORT Progress Report,  
Jul. - Sep. 1986 (Jet Propulsion Lab.)  
313 p

CSCL 17B

N87-15329  
THRU  
N87-15360  
Unclassified  
40128

G3/32

# **The Telecommunications and Data Acquisition Progress Report 42-87**

## **July – September 1986**

**E. C. Posner**  
Editor

November 15, 1986



National Aeronautics and  
Space Administration

**Jet Propulsion Laboratory**  
California Institute of Technology  
Pasadena, California

The research described in this publication was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.

## Preface

This quarterly publication provides archival reports on developments in programs managed by JPL's Office of Telecommunications and Data Acquisition (TDA). In space communications, radio navigation, radio science, and ground-based radio astronomy, it reports on activities of the Deep Space Network (DSN) and its associated Ground Communications Facility (GCF) in planning, in supporting research and technology, in implementation, and in operations. Also included is TDA-funded activity at JPL on data and information systems and reimbursable DSN work performed for other space agencies through NASA. The preceding work is all performed for NASA's Office of Space Tracking and Data Systems (OSTDS).

In geodynamics, the publication reports on the application of radio interferometry at microwave frequencies for geodynamic measurements. In the search for extraterrestrial intelligence (SETI), it reports on implementation and operations for searching the microwave spectrum. The latter two programs are performed for NASA's Office of Space Science and Applications (OSSA).

Finally, tasks funded under the JPL Director's Discretionary Fund and the Caltech President's Fund which involve the TDA Office are included.

This and each succeeding issue of the TDA Progress Report will present material in some, but not necessarily all, of the following categories:

### OSTDS Tasks:

- DSN Advanced Systems
  - Tracking and Ground-Based Navigation
  - Communications, Spacecraft-Ground
  - Station Control and System Technology
  - Network Data Processing and Productivity
- DSN Systems Implementation
  - Capabilities for New Projects
  - Networks Consolidation Program
  - New Initiatives
  - Network Sustaining
- DSN Operations
  - Network Operations and Operations Support
  - Mission Interface and Support
  - TDA Program Management and Analysis
- GCF Implementation and Operations
- Data and Information Systems

### OSSA Tasks:

- Search for Extraterrestrial Intelligence
- Geodynamics
  - Geodetic Instrument Development
  - Geodynamic Science

### Discretionary Funded Tasks

## Contents

### OSTDS TASKS DSN Advanced Systems TRACKING AND GROUND-BASED NAVIGATION

<b>Reference Frequency Transmission Over Optical Fiber .....</b>	<b>1</b>
G. Lutes and A. Kirk	
NASA Code 310-10-62-16-00	

### COMMUNICATIONS, SPACECRAFT-GROUND

<b>Design Considerations for the Beam-Waveguide Retrofit of a Ground Antenna Station .....</b>	<b>10</b>
T. Veruttipong, J. Withington, V. Galindo-Israel,	
W. Imbriale, and D. Bathker	
NASA Code 310-20-65-86-02	
<b>A Synchronization Technique for Optical PPM Signals .....</b>	<b>24</b>
V. A. Vilnrotter, E. R. Rodemich, and H. H. Tan	
NASA Code 310-20-67-59-00	
<b>Received Optical Power Calculations for Optical Communications Link Performance Analysis .....</b>	<b>32</b>
W. K. Marshall and B. D. Burk	
NASA Code 310-20-67-63-00	
<b>A Cooled Avalanche Photodiode With High Photon Detection Probability .....</b>	<b>41</b>
D. L. Robinson and B. D. Metscher	
NASA Code 310-20-67-59-00	
<b>Fraunhofer Filters to Reduce Solar Background for Optical Communications .....</b>	<b>48</b>
E. L. Kerr	
NASA Code 310-20-67-59-00	
<b>A Cost-Performance Model for Ground-Based Optical Communications Receiving Telescopes .....</b>	<b>56</b>
J. R. Lesh and D. L. Robinson	
NASA Code 310-20-67-59-00	
<b>Relative Performance of 8.5-GHz and 32-GHz Telemetry Links on the Basis of Total Data Return per Pass .....</b>	<b>65</b>
M. A. Koerner	
NASA Code 310-20-64-50-00	
<b>Long-Term Amplitude and Phase Stability of the 400-kW 2.115-GHz Transmitter .....</b>	<b>81</b>
D. J. Hoppe and A. M. Bhanji	
NASA Code 310-20-64-15-00	
<b>DSN Microwave Antenna Holography Part II: Data Processing and Display of High-Resolution Effective Maps .....</b>	<b>92</b>
D. J. Rochblatt, Y. Rahmat-Samii, and J. H. Mumford	
NASA Code 310-20-65-86-06	

### STATION CONTROL AND SYSTEM TECHNOLOGY

<b>A Single VLSI Chip for Computing Syndromes in the (255, 223) Reed-Solomon Decoder .....</b>	<b>98</b>
I. S. Hsu, T. K. Truong, H. M. Shao, and L. J. Deutsch	
NASA Code 310-30-70-84-08	

PRECEDING PAGE BLANK NOT FILMED

PRECEDING PAGE BLANK NOT FILMED

<b>Table Look-Up Estimation of Signal and Noise Parameters From Quantized Observables</b>	104
V. A. Vilnrotter and E. R. Rodemich	
NASA Code 310-30-70-84-06	

<b>A Comparison of Methods for DPLL Loop Filter Design</b>	114
S. Aguirre, W. J. Hurd, R. Kumar, and J. Statman	
NASA Code 310-30-70-84-02	

<b>A Generalized Algorithm to Design Finite Field Normal Basis Multipliers</b>	125
C. C. Wang	
NASA Code 310-30-71-83-02	

## **DSN Systems Implementation CAPABILITIES FOR NEW PROJECTS**

<b>X-Band Uplink Feedcone Capabilities, Components, and Layout</b>	140
H. Marlin, A. Freiley, and R. Hartop	
NASA Code 314-30-63-01-02	

<b>A New Algorithm for Microwave Delay Estimation From Water Vapor Radiometer Data</b>	149
S. E. Robinson	
NASA Code 314-40-51-36-08	

<b>Block IIA Traveling-Wave Maser</b>	158
D. L. Trowbridge	
NASA Code 314-30-54-81-02	

<b>Antenna Servo Control System Characterization: Rate Loop Analysis for 34-m Antenna at DSS 15</b>	165
J. A. Nickerson, D. G. Cox, H. K. Smith, J. H. Engle, and H.G. Ahlstrom	
NASA Code 314-40-51-38-85	

## **ARRAYING**

<b>Software Development for the VLA-GDSCC Telemetry Array Project</b>	174
H. W. Cooper and L. R. Hileman	
NASA Code 314-40-56-09-05	

<b>Real-Time Combiner Loss</b>	179
M. K. Simon and A. Mileant	
NASA Code 314-40-56-09-02	

## **NETWORK SUSTAINING**

<b>PPM/NAR 8.4-GHz Noise Temperature Statistics for DSN 64-Meter Antennas, 1982-1984</b>	185
S. D. Slobin and E. M. Andres	
NASA Code 314-40-21-30-04	

<b>Riveted Panel Surface Measurement Using Photogrammetry</b>	193
W. D. Merrick, V. B. Lobb, F. L. Lansing, and F. W. Stoller	
NASA Code 314-40-44-14-19	

<b>Pointing Calibration of the MKIVA DSN Antennas Voyager 2 Uranus Encounter Operations Support</b>	206
R. Stevens, R. L. Riggs, and B. Wood	
NASA Code 314-40-41-81-12	

## DSN Operations SUPPORT FOR HALLEY COMET MISSIONS

<b>Halley Comet Missions</b> .....	240
C. Stelzried, L. Efron, and J. Ellis	
NASA Code 314-08-10-21-40	
 <b>Giotto Mission Support</b> .....	243
C. Stelzried and T. Howe	
NASA Code 314-08-10-21-40	
 <b>Giotto Navigation Support</b> .....	249
N. A. Mottinger and R. I. Premkumar	
NASA Code 314-08-10-21-70	
 <b>Pathfinder Operations</b> .....	263
J. Wilcher, C. Stelzried, and S. Finley	
NASA Code 314-40-22-39-05	
 <b>VEGA Pathfinder Navigation for Giotto Halley Encounter</b> .....	268
J. Ellis and T. P. McElrath	
NASA Code 314-40-22-42-01	
 <b>Selection of Radio Sources for Venus Balloon-Pathfinder ΔDOR Navigation at 1.7 GHz</b> .....	279
K. M. Liewer	
NASA Code 314-40-22-12-10	
 <b>ICE Second Halley Radial: TDA Mission Support and DSN Operations</b> .....	285
N. A. Fanelli, L. Efron, and R. J. Muellerschoen	
NASA Code 314-40-22-42-01	

## MISSION INTERFACE AND SUPPORT

<b>Propagation Through Martian Dust at 8.5 and 32 GHz</b> .....	291
E. K. Smith and W. L. Flock	
NASA Code 314-40-22-70-12	

## DISCRETIONARY FUNDED TASKS

<b>Systolic VLSI for Kalman Filters</b> .....	300
H.-G. Yeh and J. J. Chang	
NASA Code 404-00-73-02-87	

# Reference Frequency Transmission Over Optical Fiber

G. Lutes and A. Kirk

Time and Frequency Systems Research Group

A 100-MHz reference frequency from a hydrogen maser frequency standard has been transmitted via optical fiber over a 14-km distance with a measured stability of  $1.5 \times 10^{-15}$  for 1000 seconds averaging time. This capability was demonstrated in a frequency distribution experiment performed in April, 1986. The reference frequency was transmitted over a single-mode fiber-optic link from Deep Space Station (DSS) 13 to DSS 12 and back to DSS 13.

This article will discuss the background leading up to the experiment and the significance of stable reference frequency distribution in the Deep Space Network (DSN). It will also describe the experiment, including the fiber-optic link, the measurement method and equipment, and finally the results of the experiment.

## I. Introduction

The Deep Space Network (DSN) has very stringent requirements for precise frequency and time. JPL uses hydrogen maser frequency standards in the DSN to meet these requirements. They are currently capable of stability performance at the level of better than one part in  $10^{-15}$  over averaging intervals of 1000 seconds. The future goal for frequency stability in the DSN requires improvements of up to two orders of magnitude over the present capability.

Such stringent requirements demand the development of a new generation of ultrastable sources of frequency, such as heavy ion frequency standards. The performance level of these new standards will place extreme stability requirements on the transmission systems that must distribute these reference frequencies to the users.

Group delay variations occur in all cables and make ultrastable frequency distribution difficult. As the group delay changes in a cable it causes the frequency of a signal passing through the cable to be offset by an amount proportional to the rate of change. Temperature changes, atmospheric pressure changes, and bending of a cable produce these group delay changes in the cable that degrade the stability of a reference frequency transmitted through the cable. Temperature changes have, by far, the greatest effect on group delay and, therefore, cause most of the degradation in frequency stability of a signal passing through a cable.

Currently, coaxial cables are used to distribute precise reference frequencies from a single hydrogen maser within a station to a number of local users. It is clear, however, that the distribution of precise frequency and time from a centralized

location to remote users within an entire complex offers some attractive benefits. Such benefits include the possibility of arraying stations for experiments such as connected element interferometry, an economy in the number of frequency standards within the complex, and a redundancy in frequency standards. This latter feature will guarantee the availability of hydrogen masers during critical periods such as encounters, and will provide for the capability of continuous characterization of the performance of the frequency source.

JPL has recognized the advantages of an ultrastable distribution system for the past several years, and efforts have been underway to develop a technology to enable a centralized frequency and timing facility. Early in the inception of this concept it was recognized that optical fiber was the most effective media for this application. This is primarily because optical fiber has superior qualities with respect to stability, immunity to radio frequency interference (RFI) and electromagnetic interference (EMI) and ability to transmit high-frequency signals over tens of kilometers without repeaters.

JPL initiated development of fiber-optic distribution systems for precise time and frequency as a task in 1979. The initial efforts included a survey of the technological trends, acquisition and evaluation of components, and the determination of delay stability and cable losses. In the following years the scope of this effort expanded to include the installation of a 3-km experimental link at JPL, and the demonstration of analog signal and digital data transmission. Other accomplishments in the effort included the demonstration of frequency division multiplexing and two-way transmission in the same fiber.

Following this work, the effort was directed to the use of a negative feedback technique to stabilize the delay through the fiber (Refs. 2, 3). This resulted in a stability of a few parts in  $10^{-16}$  for 1000 seconds averaging time over the stabilized 3-km link at JPL.

Transmitters, isolators, and techniques for 1300-nm systems were developed next and led to the installation of a cable to link DSS 12 and DSS 13 in the Goldstone Deep Space Communications Complex. This cable contains four multimode fibers, which were spliced and tested in 1983, and two experimental single-mode fibers. The multimode fiber in the 7-km cable became the first optical fiber link in the DSN, and has been used to demonstrate the replacement of microwave links between stations for the transmission of data, voice, television and time references.

After the installation of the single-mode fibers in the cable was completed, the magnitude of changes in group delay and rate of change of group delay in the cable had to be deter-

mined so an optimum electronic stabilizer, to be used to further stabilize the link, could be designed. The purpose of the experiment reported here was to determine these parameters and to measure the frequency stability of a signal passing through the unstabilized cable. The stability was measured with a frequency stability analyzer that was designed and fabricated at JPL.

After the measurements were completed the link was set up for one-way distribution of the signal from DSS 13 to DSS 12 to be used as the station reference for a planned coherent interferometry experiment.

## II. Fiber-Optic System

A fiber-optic cable containing 6 fibers was buried between Deep Space Station (DSS) 13 and DSS 12, a distance of approximately 7 km. Two of the fibers in this cable are the first single-mode fibers manufactured by Corning Glass Works to be used in a nonmilitary installation. They have an 8.5- $\mu\text{m}$  core diameter, 125- $\mu\text{m}$  cladding diameter and are designed to operate at a 1300-nm wavelength. The experiment was performed on these fibers.

One of the first concerns in trying to achieve ultrastable frequency transmission through a cable is to ensure that the delay instabilities in the cable are minimized as much as is practical by such passive means as the location and the method of cable installation before electronic stabilizing systems are employed.

In order to minimize temperature variations, the cable was buried to a depth of 1.5 m, as deep as was practical with commercial equipment. The cable was plowed directly into the ground with a nonvibrating cable plow after the ground was prepped. This was to keep stresses on the cable to a minimum. Fusion splicing was used to avoid splice reflections that can reduce the effectiveness of a cable stabilizer.

The total loss in each of the two single-mode fibers is approximately 14 dB. This includes splice loss and a loss of 3 dB in a directional coupler at one end of the fiber. It also includes mismatch losses, due to differently sized fiber cores in the fibers in the cable and the fibers in the pigtails which have factory installed connectors.

A good fiber-optic cable, such as this one, has a temperature coefficient of delay of about 7 ppm/K (Ref. 1). The sensitivity to atmospheric pressure change has not been directly measured but appears to be quite small.

A commercial fiber-optic transmitter designed to transmit television signals was used. We modified the transmitter by

bypassing the filters and DC restorer circuit to make it operate with a 100-MHz sinewave. The transmitter uses a 1300-nm single-mode semiconductor laser that emits 0.5 mW of optical power into the optical fiber. It contains electronic circuitry to control the optical output power of the laser and to keep its temperature constant.

The receiver was designed and fabricated by personnel in the Time and Frequency Systems Research Group. It has 400-MHz bandwidth which results in good temperature stability, and long-term phase stability. It remains linear under the large signal conditions encountered in this application. This provides the maximum possible signal-to-noise ratio, and therefore the best short term phase noise.

A schematic of the receiver is shown in Fig. 1. It contains a PIN photodiode, a bias circuit, a decoupling circuit and a multimode fiber pigtail. The optical input is coupled to the PIN photodiode through a multimode optical fiber. The PIN photodiode is an Indium Gallium Arsenide (InGaAs) device. Biasing is provided by a low noise voltage source (Ref. 4) consisting of a constant current diode and a low noise zener diode. Two wideband transformers loaded with their characteristic impedances provide wideband high impedance decoupling to the photodiode. The output is coupled through an SMA RF connector to a low noise wideband amplifier with 50-ohms input impedance.

### III. Fiber-Optic System Configuration

Figure 2 shows the fiber-optic link configuration used for the experiment. An RF power divider splits the RF reference signal from the hydrogen maser frequency standard, at DSS 13, into two signals. One of the signals is connected to the fiber-optic transmitter and the other signal is connected to the reference input to the phase stability analyzer. The signal going to the fiber-optic transmitter modulates the optical carrier and the resulting optical signal is transmitted to DSS 12, where it is connected to the other fiber, and returned to DSS 13. At DSS 13 the optical signal is detected to recover the 100-MHz RF signal, which is amplified to the required level and then connected to one of the other ports of the stability analyzer.

### IV. Stability Analyzer

A block diagram of the stability analyzer is shown in Fig. 3. A 100-MHz reference signal derived from the hydrogen maser frequency standard is applied to the analyzer through an isolation amplifier. An RF power splitter splits the 100-MHz reference signal into two equal amplitude signals. An offset synthesizer offsets one of these signals by 1 Hz resulting in a

frequency of 99,999,999 Hz. An RF mixer multiplies the 99,999,999-Hz signal and the 100-MHz reference signal and produces the difference frequency of 1 Hz, plus the phase noise of the signal being measured. This 1-Hz output signal passes through a low pass filter, which has gain, to a zero crossing detector. The signal out of the zero crossing detector is analyzed in a microprocessor which provides Allan variance data at 1, 2, 4, and  $8 \times 10^n$  seconds, where  $n$  is an integer  $\geq 0$ .

### V. Measurement Method

The stability analyzer measured the frequency stability (Allan variance) of the output signal from the round-trip fiber-optic link, using the input signal to the fiber-optic link as the reference. The phase difference between the two signals was also monitored and recorded.

Figure 6 shows a simplified block diagram of the test setup. For this test all the terminal equipment and test instrumentation was located at DSS 13 and a second fiber, identical to the first, was used to complete the 14-km round-trip between DSS 13 to DSS 12 and back to DSS 13. It is important to note that the data gathered was for the entire round-trip including terminal equipment.

The noise floor of the measurement equipment was determined by bypassing the entire fiber-optic link and using a coaxial attenuator in its place to maintain identical signal levels to the test equipment. Figure 7 shows the resulting Allan variance data which is marked "N.F." (noise floor). Notice the deviation from a  $1/\tau$  slope where  $\tau > 10$  seconds. This is primarily due to thermal and vibrational characteristics of the test environment on the stability analyzer and test cables.

An existing clean-up loop with a noise bandwidth of 5 Hz was used at the output of the fiber-optic link to reduce the bandwidth and measured noise power of the received signal. The noise floor of the test system, with the clean-up loop only, was measured and is shown in Fig. 7 and is marked "C.U.L." (clean-up loop).

An additional measurement was performed using a short piece of fiber-optic cable at the test location in order to determine the stability of the terminal equipment only. This is shown in Fig. 7 by "T.E." (terminal equipment).

### VI. Test Results

Figure 8 shows the Allan variance for the roundtrip fiber-optic link, including the clean-up loop. The  $1/\tau$  slope portion of the graph where  $\tau$  is less than 100 s is primarily due to the

signal-to-noise ratio of the fiber-optic link. For  $\tau > 100$  s the Allan variance does not decrease at a  $1/\tau$  slope as expected but shows a hump around 400-s averaging time.

The frequency stability of signals passing through cables which are subjected to temperature cycling is expected to be degraded. This degradation shows up on Allan variance plots as a hump usually between 100 and 1000 seconds which corresponds to the air conditioning cycling time. The time constant of the cable, the length of the exposed cable, and the magnitude of the temperature variation determine the magnitude of the hump. The hump is usually broad because the cycling time is not constant.

The cyclic phase change responsible for the hump in the Allan variance curve was observed and recorded. It appeared to be a result of temperature changes in the plenum at DSS 12. The air conditioning at this location has a cycle period of about 1080 seconds, and the peak to peak temperature variation is about 3 K. The length of the exposed cable in the plenum is about 45.7 m round-trip and the cable has a temperature coefficient of 7 ppm/K.

The expected phase variation as a result of this temperature change was calculated using the above values and, assuming zero time constant for the cable, the phase change in degrees as a result of a change in temperature is,

$$\theta = 1.71 \times 10^{-12} L \alpha T f_o \quad (1)$$

where

$L$  = the affected length of cable, m

$\alpha$  = the temperature coefficient of delay of the cable, ppm/K

$T$  = the change in temperature, K

$f_o$  = the operating frequency

Evaluating this equation using the above values yields 0.164 deg, which is about twice the variation measured. However, the integration of the temperature variation as a result of the time constant of the cable could account for the difference. The time constant of this cable was calculated from this data as 5 min. This is not unreasonable since a coaxial cable of approximately the same size, but with more mass, has a time constant of  $\approx 20$  min.

To verify that this effect was actually observed, the phase of the round-trip signal was compared to the reference signal and monitored at DSS 13. The temperature in the plenum at

DSS 12 was also monitored at the same time. A strong correlation was observed between the variation of temperature and the variation of phase. In the evening, when the outside temperature is lower, the air conditioning in the DSS 12 plenum is turned off and outside air is used to cool the building. During this period the temperature did not cycle and the observed phase variations at DSS 13 disappeared. This was a further indication that the observed phase variations at DSS 13 were a result of the temperature variations in the plenum at DSS 12. The plots of phase at DSS 13 and temperature at DSS 12 are shown in Figs. 4 and 5, respectively.

As mentioned before, the link was set up for one-way transmission of the hydrogen maser reference signal from DSS 13 to DSS 12 to provide a stable reference for DSS 12. The stability of the one-way trip cannot be measured because there is no independent reference at the far end of the link. However, it can be estimated for the  $1/\tau$  region of the curve from the relationship between the Allan variance and the signal-to-noise ratio of the signal being measured (Ref. 6).

$$\sigma = \frac{1}{\omega\tau} \sqrt{\frac{P_n}{P_s}} \quad (2)$$

where

$\omega$  = the angular operating frequency

$\tau$  = the sampling interval

$P_n$  = the noise power in the signal being measured

$P_s$  = the signal power

Using the SNRs measured at DSS 12 and DSS 13 (110 dB/Hz and 85 dB/Hz, respectively) and the above relationship we estimate a theoretical improvement of 25 dB for the one-way trip for averaging times less than 10 seconds. This improvement will not be fully realized because the cleanup loop has a bandwidth of 5 Hz corresponding to a stability of only  $1 \times 10^{-13}$  for 1 second averaging times. A cleanup loop having the proper bandwidth is being designed.

It was also noticed that bending the fiber resulted in phase changes of several degrees. This was unexpected based on experiments performed earlier and the results of a literature search.

## VII. Discussion

The experiment reported here has generated a number of new and interesting questions. Important questions pertain-

to why the delay through the fiber-optic link is affected so much by bending the fiber. According to the theory (Ref. 5) the change in delay of the fiber should be much smaller. One may speculate that the amplitude to phase modulation conversion in the measurement system could be responsible for the change in the delay, or it could be that changing reflections back into the laser changes its wavelength, resulting in delay variations due to dispersion in the fiber. A third possibility is birefringence in the fiber. However, the most probable cause of this group delay change is the optical equivalent of voltage-standing-wave ratio. We are currently engaged in the study of the effect to identify the mechanism responsible for it and ways to minimize it.

How stable is the buried cable? The signal-to-noise ratio of the laser, temperature variations, and floor vibrations appear to account for most of the frequency instability we measured. If this is true, we have not yet seen the noise of the buried fiber-optic cable and we must reduce these instabilities in order to see the cable instabilities. Increasing the time constant of the exposed fiber-optic cable and isolating the

equipment from vibrations should improve the frequency stability of the distribution system.

## VIII. Summary

A 100-MHz reference frequency from a hydrogen maser has been distributed 14 km over an unstabilized single-mode fiber-optic link with a stability of  $1.5 \times 10^{-15}$  for 1000 seconds averaging time. This link is providing an ultrastable reference signal to DSS 12 from DSS 13 over a distance of 7 km. This reference is more than 2 orders of magnitude better than the station Cesium beam reference normally used at DSS 12. This will permit coherent interferometry measurements to be made using the two stations.

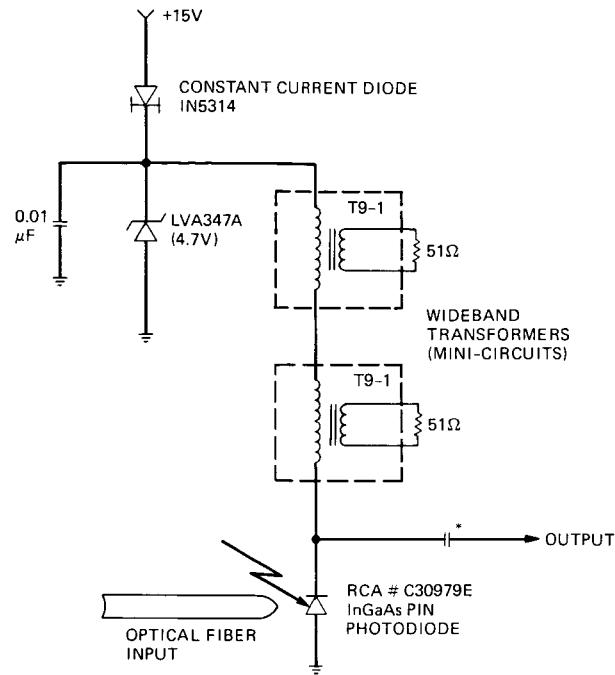
This experiment has provided the data necessary to greatly improve the stability of reference frequency distribution systems; however, there is still much work to be done to achieve the ultimate goal of  $10^{-18}$  stability for 1000 seconds averaging time.

## Acknowledgment

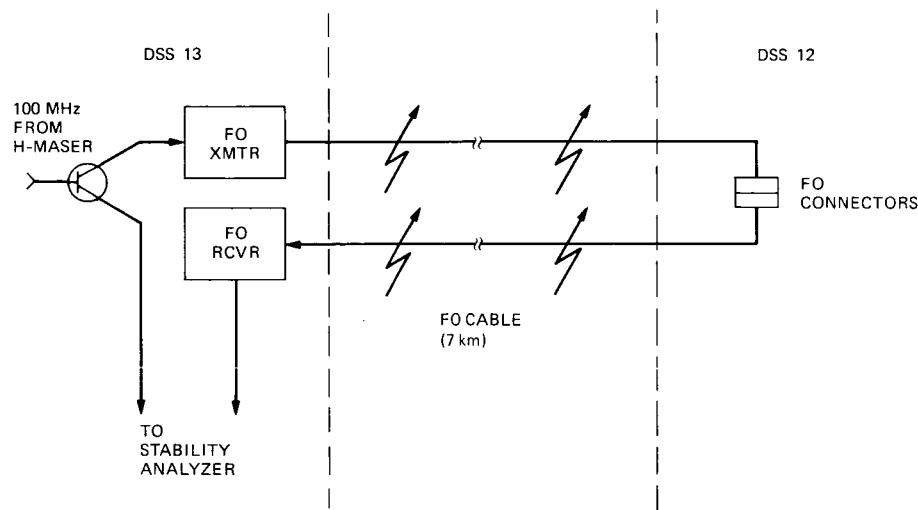
The authors wish to thank Dr. R. Sydnor, Dr. L. Maleki, Dr. C. Greenhall, T. Tucker, R. Meyer, W. Diener, L. Patterson and P. Tu for their suggestions and help during the course of this work. We would also like to thank B. Walker, F. Sullens, E. Crofoot and all of the other communications and facilities personnel at Goldstone for their help and cooperation.

## References

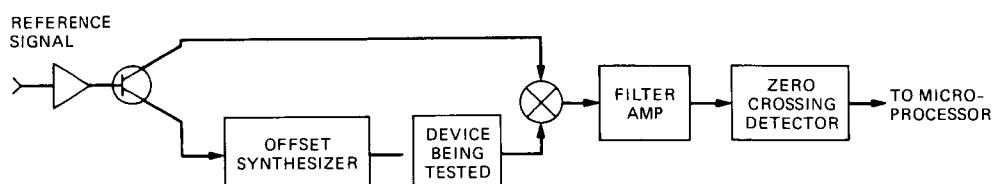
1. Bergman, L. A., Eng, S. T., Johnston, A. R., and Lutes, G. F., "Temperature dependence of phase for a single-mode fiber cable," *Proceedings of Third International Conference on Integrated Optics and Optical Fiber Communications*, p. 60, OSA-IEEE, April 27-29, 1981, San Francisco, CA.
2. Lutes, G., "Optical fibers for the distribution of frequency and timing references," *Proceedings of the 12th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, pp. 663-680, NASA Conference Publication 2175, Goddard Space Flight Center, Dec. 1980.
3. Lutes, G., "Development of optical fiber frequency and time distribution systems," *Proceedings of the 13th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, pp. 243-262, NASA Conference Publication 2220, Naval Research Laboratory, Dec. 1981.
4. Lutes, G., "Stable low noise voltage source," *DSN Progress Report 42-47*, pp. 89-93, Jet Propulsion Laboratory, Pasadena, CA, July and August 1978.
5. Lau, K. Y., "Propagation path length variations due to bending of optical fibers," *The Telecommunications and Data Acquisition Progress Report 42-63*, pp. 26-32, Jet Propulsion Laboratory, Pasadena, CA, March-April, 1981.
6. Cutler, L. S., and Searle, C. L., "Some Aspects of the Theory and Measurement of Frequency Fluctuations in Frequency Standards," *Proceedings of the IEEE*, special issue on Frequency Stability, Vol. 54, pp. 136-154, February, 1966.



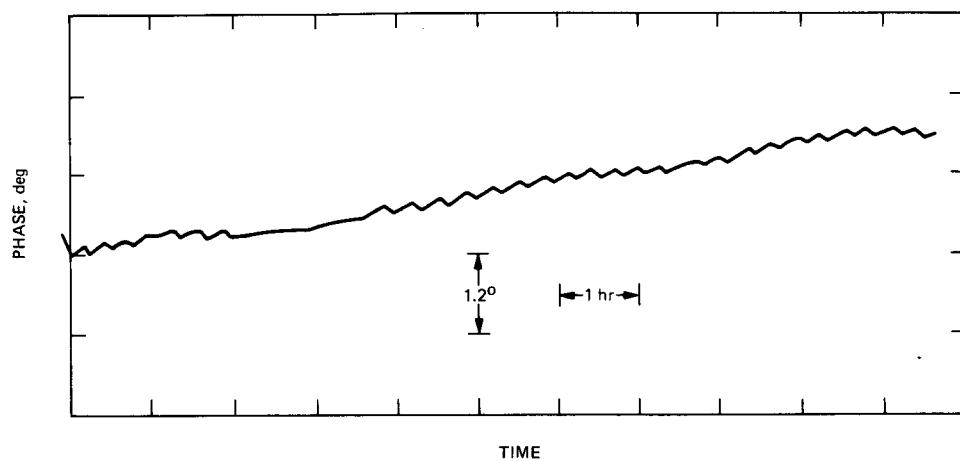
**Fig. 1. Fiber-optic receiver schematic**



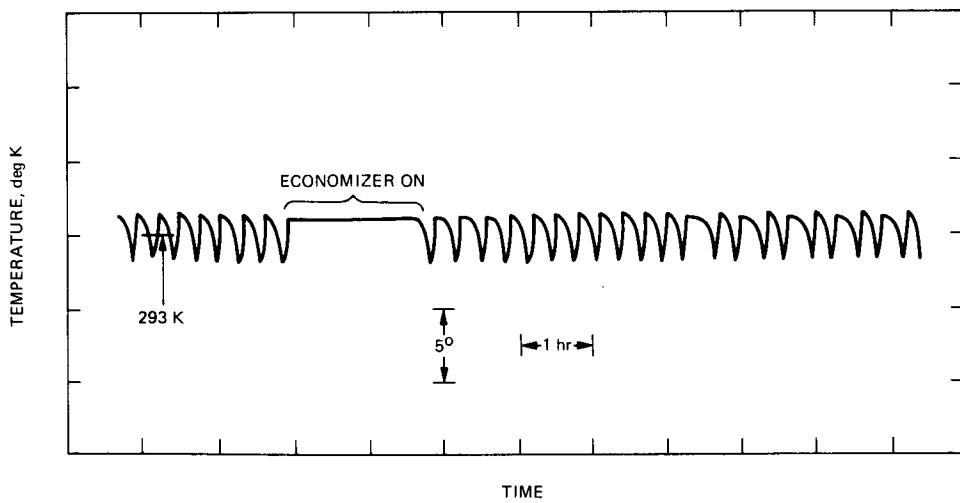
**Fig. 2. Fiber-optic (FO) system configuration**



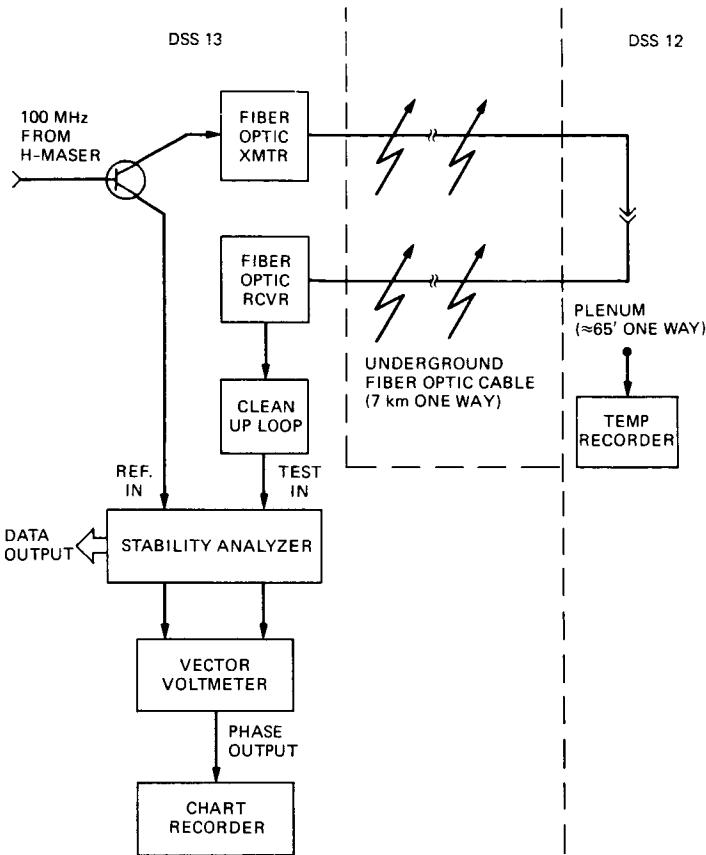
**Fig. 3. Simplified block diagram of the stability analyzer**



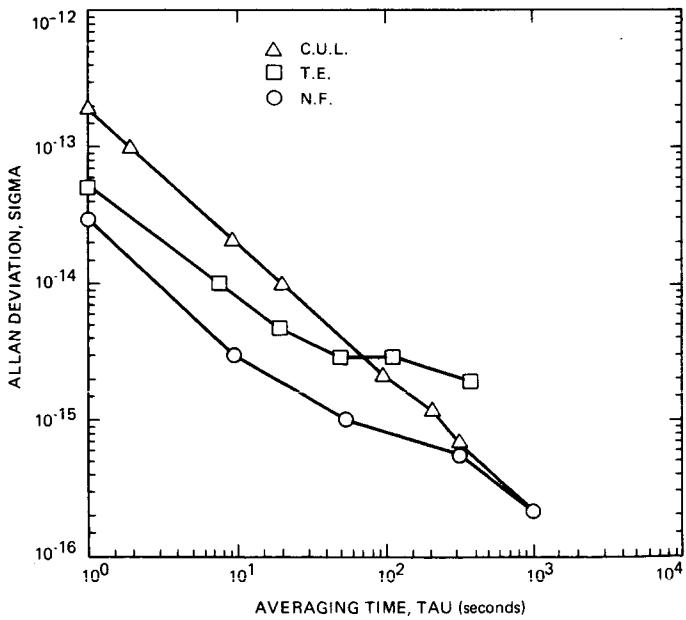
**Fig. 4. Phase measured across the fiber-optic link**



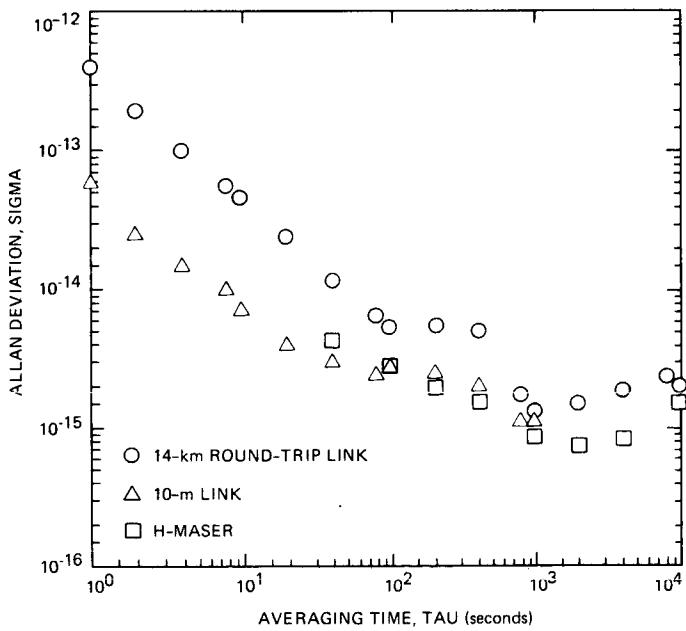
**Fig. 5. Temperature measured in the Echo station plenum**



**Fig. 6. Simplified block diagram of the test set-up**



**Fig. 7. Allan variance vs averaging time for the cleanup loop (C.U.L.), terminal equipment (T.E.), and noise floor (N.F.)**



**Fig. 8. Allan variance vs averaging time for the 14-km link, 10-m link, and hydrogen maser**

# Design Considerations for the Beam-Waveguide Retrofit of a Ground Antenna Station

T. Veruttipong, J. Withington, V. Galindo-Israel,  
W. Imbriale, and D. Bathker  
Radio Frequency and Microwave Subsystems Section

*Retrofitting an antenna that was originally designed without a beam waveguide introduces special difficulties because it is desirable to minimize alteration of the original mechanical truss work and to image the actual feed without distortion at the focal point of the dual-shaped reflector. To obtain an acceptable image, certain Geometrical Optics (GO) design criteria are followed as closely as possible. The problems associated with applying these design criteria to a 34-meter dual-shaped DSN antenna are discussed. The use of various diffraction analysis techniques in the design process is also discussed. GTD and FFT algorithms are particularly necessary at the higher frequencies, while Physical Optics and Spherical Wave Expansions proved necessary at the lower frequencies.*

## I. Introduction

A primary requirement of the NASA Deep Space Network (DSN) is to provide for optimal reception of very low signal levels. This requirement necessitates optimizing the antenna gain to the total system operating noise level quotient. Low overall system noise levels of 16 to 20 K are achieved by using cryogenically cooled preamplifiers closely coupled with an appropriately balanced antenna gain/spillover design. Additionally, high-power transmitters (up to 400 kW CW) are required for spacecraft emergency command and planetary radar experiments. The frequency bands allocated for deep space telemetry are narrow bands near 2.1 and 2.3 GHz (S-band), 7.1 and 8.4 GHz (X-band), and 32 and 34.5 GHz (Ka-band). In addition, planned operations for the Search for Extraterrestrial Intelligence (SETI) program require continu-

ous low-noise receive coverage over the 1 to 10 GHz band. To summarize, DSN antennas must operate efficiently with low receive noise and high-power uplink over the 1 to 35 GHz band.

Feeding a large low-noise, ground-based antenna via a beam-waveguide system has several advantages over directly placing the feed at the focal point of a dual-shaped antenna. For example, significant simplifications are possible in the design of high-power, water-cooled transmitters and low-noise cryogenic amplifiers, since these systems do not have to rotate as in a normally fed dual reflector. Furthermore, these systems and other components can be placed in a more accessible location, leading to improved service and availability. Also, the losses associated with rain on the feedhorn radome

are eliminated because the feedhorn can be sheltered from weather.

Many existing beam-waveguide systems use a quasi-optical design, based on Gaussian wave principles, which optimizes performance over an intended operating frequency range. These designs can be made to work well with relatively small reflectors (a very few tens of wavelengths), and may be viewed as "bandpass," since performance suffers as the wavelength becomes very short as well as very long. The long wavelength end is naturally limited by the approaching small  $D/\lambda$  of the individual beam reflectors used; the short wavelength end does not produce the proper focusing needed to image the feed at the dual-reflector focus. In contrast, a purely geometrical optics (G.O.) design has no upper frequency limit, but performance suffers at long wavelengths. These designs may be viewed as "high pass." Considering the need for practically sized beam reflectors and the high DSN frequency and performance requirements, the G.O. design is favored in this application.

Retrofitting an antenna that was originally designed without a beam waveguide introduces special difficulties because it is desirable to minimize alteration of the original structure. This may preclude accessing the center region of the reflector (typically used in conventional beam-waveguide designs), and may require bypassing the center region. A discussion of the mechanical tradeoffs and constraints is given herein, along with a performance analysis of some typical designs. In the retrofit design, it is also desirable to image the original feed without distortion at the focal point of the dual-shaped reflector. This will minimize gain loss, reflector design, and feed changes.

To obtain an acceptable image, certain design criteria are followed as closely as possible. In 1973, Mizusawa and Kitamura (Ref. 1) introduced certain G.O. criteria which guarantee a perfect image from a reflector pair (cell). If more than one cell is used (where each cell may or may not satisfy Mizusawa's criteria), application of other G.O. symmetry conditions can also guarantee a perfect image. The problems and opportunities associated with applying these conditions to a 34-m dual-shaped antenna are discussed.

The use of various diffraction analysis techniques in the design process is also discussed. Gaussian (Goubau) modes provide important insight to the wave propagation characteristics, but Geometrical Theory of Diffraction (GTD), FFT, Spherical Wave Expansion (SWE), and Physical Optics (PO) have proven more accurate and faster. The GTD and FFT algorithms are particularly necessary at the higher frequencies. Both PO and SWE have been necessary at the lower frequencies.

## II. Design Considerations

### A. High-Pass Design Feed Imaging

For a 2.4-m (8-foot) reflector beam waveguide operated over 1 to 35 GHz with near perfect imaging at X-band and Ka-band and acceptable performance degradation at L-band (1.68 GHz) and S-band, a high-pass type beam waveguide should be used. This type of design is based upon G. O. and Mizusawa's criteria. Mizusawa's criteria can be briefly stated as follows:

For a circularly symmetric input beam, the conditions on a conic reflector pair necessary to produce an identical output beam are:

- (1) The four loci (two of which may be coincident) associated with the two curved reflectors must be arranged on a straight line.
- (2) The eccentricity of the second reflector must be equal to the eccentricity or the reciprocal of the eccentricity of the first reflector.

Figure 1 shows some of the orientations of the curved reflector pair that satisfy Mizusawa's criteria. We term a curved reflector pair as one cell.

For the case of two cells where at least one cell does not satisfy Mizusawa's criteria, a perfect image may still be achieved by imposing some additional conditions, described below.

Let  $S_1, S_2, S_3$ , and  $S_4$  be curved surfaces of two cells as shown in Fig. 2. Each surface can be an ellipsoid, hyperboloid, or paraboloid. Keeping the same sequence order, the surfaces are divided into two pairs [first pair ( $S_2, S_3$ ); second pair ( $S_1, S_4$ )] as shown in Fig. 2.

For a circularly symmetric input pattern, an identical output pattern (in the G.O. limit) can be obtained if Mizusawa's criteria are satisfied for both pairs in the following manner: First pair ( $S_2, S_3$ ) satisfies Mizusawa's criteria; second pair ( $S_1, S_4$ ) satisfies Mizusawa's criteria after eliminating the first pair.

It is noted that the first pair can be eliminated because the input is identical to the output. Also, this concept can be applied to cases with more than two cells. Examples of an extension of Mizusawa's criteria for a multiple-reflector beam waveguide are given in Figs. 3 and 4.

Figure 5 shows a geometrical optics field reflected from each reflector of a beam-waveguide system. (Refer to Fig. 3[a].) It is clear from Fig. 5 that the distorted pattern from the first

cell is completely compensated for by the second cell and yields an output pattern identical to the input pattern.

## B. Bandpass Design Feed Imaging

For many systems, a single-frequency or bandpass design can be advantageously employed. The design considerations can best be described with reference to Fig. 6, where the center frequency is given as  $f_0$ , and  $L_2$  is the spacing between two curved surfaces. A bandpass beam-waveguide system is usually composed of two non-confocal (shallow) ellipsoids (eccentricity close to one). Again from Fig. 6,  $F_{A1}$  and  $F_{A2}$  are G.O. foci of ellipsoid  $A$ , while  $F_{AP}$  is the phase center of the scattered field from surface  $A$  (evaluated at frequency =  $f_0$ ) in the neighborhood of surface  $B$ . Similarly,  $F_{B1}$ ,  $F_{B2}$ , and  $F_{BP}$  are for ellipsoid  $B$ . The distances from  $F_{A2}$  and  $F_{AP}$  to surface  $A$  are very large compared to  $L_1$  and  $L_2$ . The locations of  $F_{AP}$  and  $F_{BP}$  depend on frequency as well as surface curvature,  $L_1$ , and  $L_2$ . For example, with a 2.4-m reflector with eccentricity = 0.97 at  $f = 2.3$  GHz and  $L_2 = 8$  m (26 feet),  $F_{AP}$  is about 120 m (400 feet) to the left of ellipsoid  $A$ , as shown in Fig. 6(a). In the G.O. limit,  $F_{AP}$  and  $F_{A2}$  are at the same location, to the right of ellipsoid  $A$ .

It is desirable to have two identical surfaces for low cross-polarization and a symmetrical system. Trial and error are needed in order to determine surface parameters for the desired operating frequency and bandwidth within specified losses.

Figure 7 shows the input and output patterns from a bandpass beam-waveguide system where  $F_{AP}$  and  $F_{BP}$  are chosen to be at the same locations as  $F_{B2}$  and  $F_{A2}$ , respectively (the choice may not be the optimum condition). The two identical ellipsoids are designed at  $f_0 = 2.3$  GHz. The results show good agreement between the input feed and the imaged feed. Bandpass beam-waveguide systems appear useful when a limited band coverage is required, using modestly sized ( $D = 20$  to  $30\lambda$ ) reflectors. However, these systems do not perform well as the wavelength approaches either zero or infinity. In contrast, a high-pass (G.O.) design focuses perfectly at zero wavelength and focuses very well down to  $D \sim 40\lambda$ . The performance then decreases monotonically as  $D$  becomes smaller in wavelength.

## III. Application Considerations for the DSN

The DSN presently operates three 34-m high efficiency (H.E.) dual-shaped reflector antennas with a dual-band (2.3/8.4-GHz) feed having a far-field gain of +22.4 dBi that is conventionally located at the Cassegrain focal point. The structures were designed prior to the beam-waveguide requirements, and feature a continuous elevation axle and a carefully designed

elevation wheel substructure. The elevation wheel substructure, shown in Fig. 8, plays a key role in preserving main reflector contour integrity as the antenna rotates in elevation. To maintain contour integrity at 8 GHz and above is of prime concern for RF efficiency performance as well as retrofit costs. Figure 9 shows a centerline beam-waveguide approach which, although it is a compact and straightforward RF design, severely impacts retrofit costs and the contour integrity of the main reflector (hence RF efficiency at 8 GHz and above). Figure 10 shows an unconventional approach and represents attempts to reduce structural impacts. Figure 10 is a six-reflector beam waveguide (eight-reflector antenna) based on our extension of Mizusawa's criteria. Although several detailed options are possible, most options use two cells (four curved reflectors) with two flat reflectors. Some of the options make use of the flexibility afforded by allowing each cell to be distorting (of itself), but then compensated for by the second cell as described in Figs. 3, 4, and 5.

The goal is therefore to perfectly image a feed located perhaps 15 to 25 m (50 to 80 feet) below the main reflector to the original Cassegrain focus. This goal applies over the 1 to 35-GHz frequency range, using a beam-waveguide housing limited to about 2.4 m (8 feet) in diameter. The image should be a 1:1 beamwidth transformation of the original +22.4 dBi feed, permitting reuse of that feed and no changes to the subreflector or main reflector contour. Lastly, the goal includes minimal structural impacts, particularly to the integrity of the main reflector contour. The generalized solution to these goals is reflected in the approach shown in Fig. 10, termed the bypass beam waveguide.

## IV. Analytical Techniques for Design and Analysis

The software requirements for the study and design of beam waveguides are extensive. These include the capability for G.O. synthesis, Gaussian wave analysis, and high and low frequency diffraction analysis. These requirements are discussed below.

### A. Geometrical Optics Synthesis Capability

This capability includes software which synthesizes as well as analyzes reflectors satisfying the Mizusawa-Kitsuregawa conditions (Ref. 1) for minimum cross-polarization and best imaging.

In the high-frequency domain (8 to 35 GHz for the designs considered herein), the focused system shown in Fig. 11 is desired. Of course, two paraboloids or mixtures of various conic-section reflectors can be synthesized. Optimization at

lower frequency bands generally is accomplished by appropriate defocusing.

### B. Gaussian Wave Analysis Capability

The required defocusing for the lower bands can be determined by using various beam imaging techniques (Ref. 2) which are based on Gaussian beam analysis methods first developed by Goubau and Schwering (Ref. 3). While Gaussian mode analysis is useful at high as well as low frequencies for "conceptual" designing, conventional diffraction analysis methods are found to be more suitable.

One consideration is that Gaussian mode analysis does not supply spillover losses directly with as great an accuracy as conventional diffraction analysis methods. The Gaussian modes supply the discrete spectrum, whereas the spillover losses are directly related to the continuous spectrum (Ref. 4, pp. 474-475). Spillover for Gaussian modes is computed by a method more suitable in terms of computational efficiency for a great number of reflection (refraction) elements (Ref. 3). The antenna systems considered here rarely contain more than four curved reflectors. Further, spillover losses must be reliably known to much less than a tenth of a decibel for high performance systems.

### C. Low Frequency Diffraction Analysis Capability

Because of the large bandwidth of operation (1 to 35 GHz), there is no single diffraction analysis method which will be both accurate and efficient over the entire band.

Efficiency, in the sense of speed of computation, is critical, since in a constrained design many different configurations may be analyzed before an optimum beam-waveguide configuration is selected.

In the region of 2.3 GHz (for the 34-m antenna considered herein), the reflector diameters are generally about  $20\lambda$ . Since a very low edge taper illumination, at least -20 dB, is used to reduce spillover loss, the "effective" reflector diameters are very small in this frequency range.

A comparison between three diffraction algorithms: PO, GTD, and Jacobi-Bessel (JB) leads to the conclusion that:

- (1) GTD is not sufficiently accurate at the low frequencies.
- (2) JB is very slowly convergent in many cases and gives only the far-field in any case. We must determine near-field patterns.
- (3) PO is both accurate and sufficiently fast below 3 GHz.

The above results are illustrated in Figs. 12 and 13 when an actual +22.4 dBi corrugated feedhorn is used. It can be seen from Fig. 13 that PO and JB agree perfectly to  $\sim \pm 20$  deg.

There are two general PO computer algorithms useful at the low frequencies. One is a straightforward PO algorithm which subdivides the reflectors into small triangular facets. This is essentially a trapezoidal integration of the near-field radiation integral and is a very flexible algorithm.

A second algorithm is based on a Spherical Wave Expansion (SWE) and is also a PO technique. It has two useful characteristics: (1) when a high degree of circular symmetry exists in the scattered fields, then the two-dimensional radiation integral is reduced to a small number of one-dimensional integrals with a resultant marked decrease of computational speed; and (2) an  $r$  interpolation of the scattered field (at different radial distances from the coordinate origin) is done very accurately.

The PO (direct-trapezoid) and SWE algorithms are useful for cross-checking results. Near-field and far-field computations and comparisons in both amplitude and phase are shown in Figs. 14, 15, and 16. Figure 14 contains near-field PO and SWE results for scattering from one paraboloid reflector. Figure 15 contains the results for far-field scattering from two reflectors. Figure 16 contains the near-field scattering for two reflectors and a comparison of the "imaging" with the feed pattern. At higher frequencies ( $> 8$  GHz), the feed pattern will be virtually perfectly imaged over an angle of  $\pm 15$  deg.

### D. High Frequency Diffraction Analysis Capability

For reflector diameters of 70 or more wavelengths ( $> 8$  GHz), including a -20 dB edge taper, PO analysis methods become too expensive and time consuming. (The SWE may still be useful if a high degree of rotational symmetry exists.) An alternative approach is to use GTD analysis. The GTD computation speed does not increase with increasing reflector diameters, but the accuracy of the analysis does increase.

In order to test the accuracy of GTD at 8 GHz, comparisons were made between GTD and PO. Results for diffraction of a single ellipsoid are shown in Fig. 17. The results for the phase of the scattered field were equally as good. GTD was determined to be accurate at 8 GHz and higher.

Analysis of two or more reflectors by GTD involves some manipulation of the fields scattered between any two reflectors. The fields scattered from one reflector must be placed in a format suitable for GTD scattering from the next reflector. This is accomplished by computing the vector-scattered field in the vicinity of the next reflector and then interpolating as follows:

- (1) Use of an FFT for  $\phi$ -variable interpolation.
- (2) Use of a second-order Lagrangian local interpolation for  $\theta$  interpolation (for a z-axis along the axis of the reflector,  $\phi$  and  $\theta$  are spherical coordinates).
- (3) For the  $r$  (of  $[r, \theta, \phi]$  spherical coordinates) interpolation, an approximation consistent with the GTD approximation was to assume a  $1/r$  variation in amplitude and a  $kr$  variation in phase. This approximate interpolation should be checked against exact computations of the near fields.

By the method described above, multiple reflector computations, even with a large number of reflectors, can be calculated with both great speed and accuracy at frequencies above  $\sim 8$  GHz for reflectors of  $> 70\lambda$  in diameter.

A typical result is shown in Fig. 18 for a pair of ellipsoids which satisfy the Mizusawa-Kitsuregawa criteria. The object is to perfectly image the input feed over about  $\pm 20$  deg. This is accomplished with great accuracy and virtually no loss at X-band frequencies and higher.

## V. Conclusions

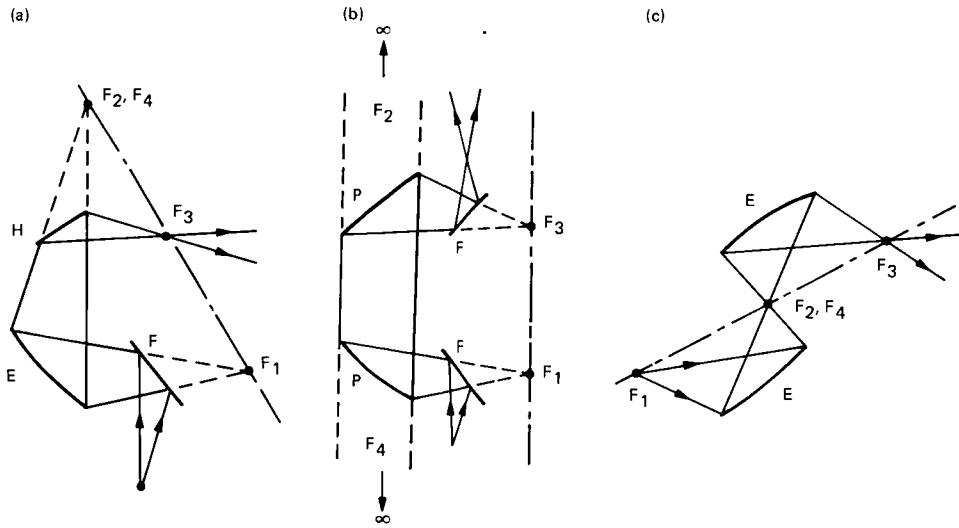
A generalized solution has been achieved for retrofitting a large dual-shaped reflector antenna for beam waveguide. The design is termed bypass beam waveguide. Several detailed options within the bypass category remain to be studied, and work continues.

With the analysis capability available, we are gaining some valuable views of the RF performance behavior of some of the many options. It appears fairly clear for the 1 to 35-GHz requirement that high-pass (pure G.O.) designs are necessary in contrast to a bandpass (non-confocal ellipsoids) approach. It appears that deep confocal ellipsoids satisfying the Mizusawa-Kitsuregawa criteria operate (focus) well with reflector diameters of about 70 and larger but may not be tolerable at longer wavelengths.

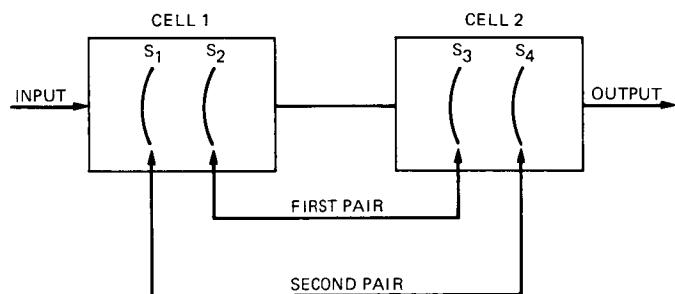
As a part of this activity, an important extension of the Mizusawa-Kitsuregawa criteria has been revealed. The principle revealed shows how a two-reflector cell, although in itself distorting, may be combined with a second cell which compensates for the first and delivers an output beam which is a good image of the input beam.

## References

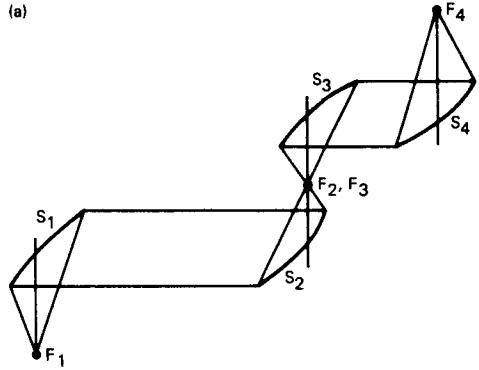
1. Mizusawa, M., and Kitsuregawa, T., A Beam Waveguide Feed Having a Symmetric Beam for Cassegrain Antennas, *IEEE Trans. Antennas and Propagat.*, pp. 844-886, November 1973.
2. Chu, T. S., An Imaging Beam Waveguide Feed, *IEEE Trans. Antennas and Propagat.*, pp. 614-619, July 1983.
3. Goubau, G., and Schwering, F., On the Guided Propagation of Electromagnetic Wave Beams, *IRE Trans. Antennas and Propagat.*, pp. 248-256, May 1961.
4. Collin, R. E., *Field Theory of Guided Waves*, McGraw-Hill, N.Y., 1960.



**Fig. 1. Possible two-curved reflector beam-waveguide configurations; E = ellipsoid, H = hyperboloid, P = paraboloid, F = flat plate**

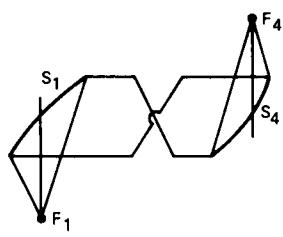


**Fig. 2. Multiple curved reflector beam-waveguide system;**  
 **$S_1$  = ellipsoid, hyperboloid, or paraboloid**

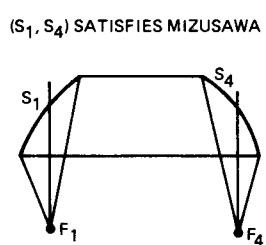


$\Downarrow$   $(S_2, S_3)$  SATISFIES MIZUSAWA  
AND IS ELIMINATED

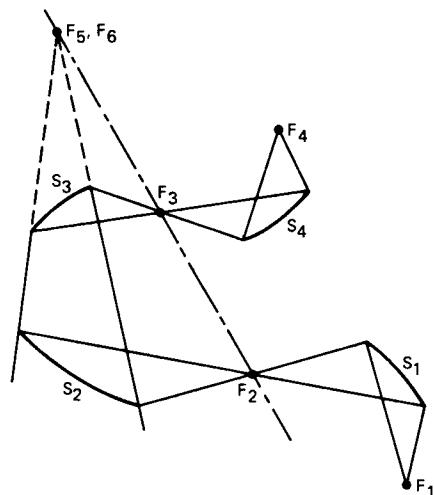
(b)



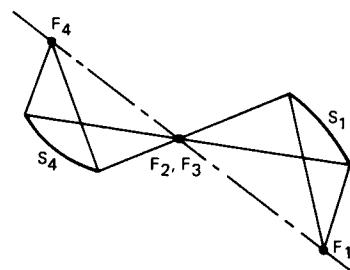
(c)



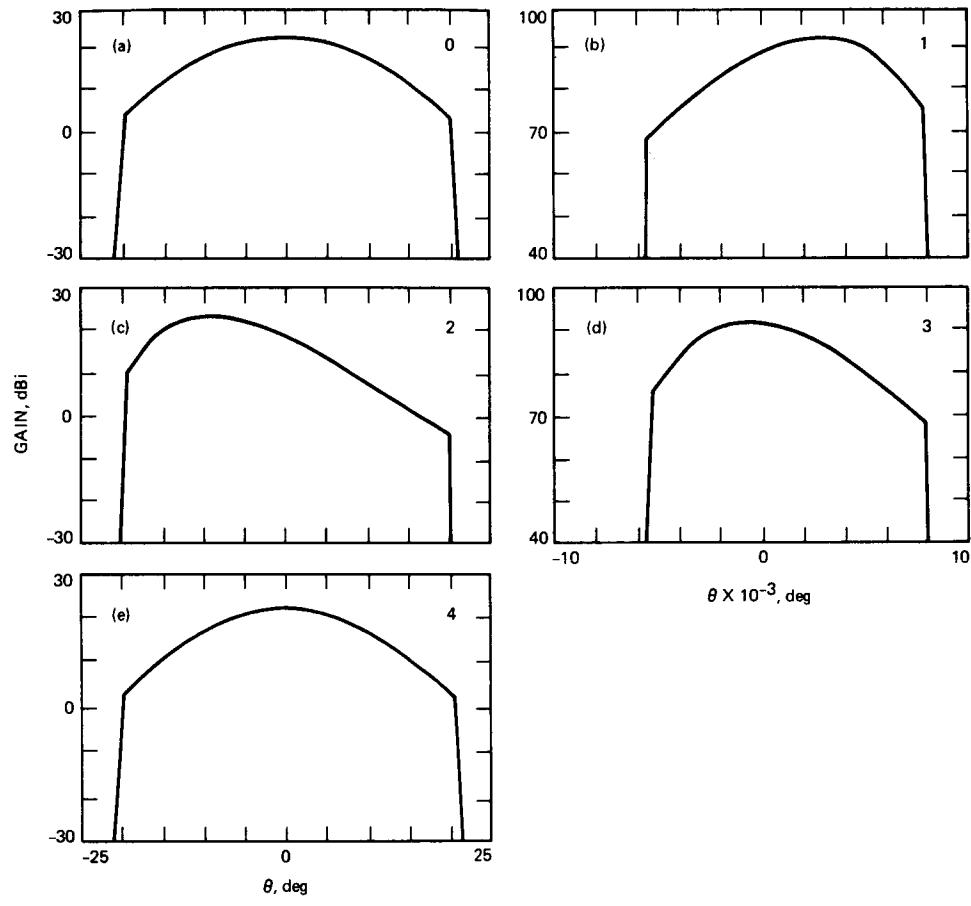
**Fig. 3. Demonstrating an extension of Mizusawa's criteria for a multiple reflector beam waveguide (paraboloids)**



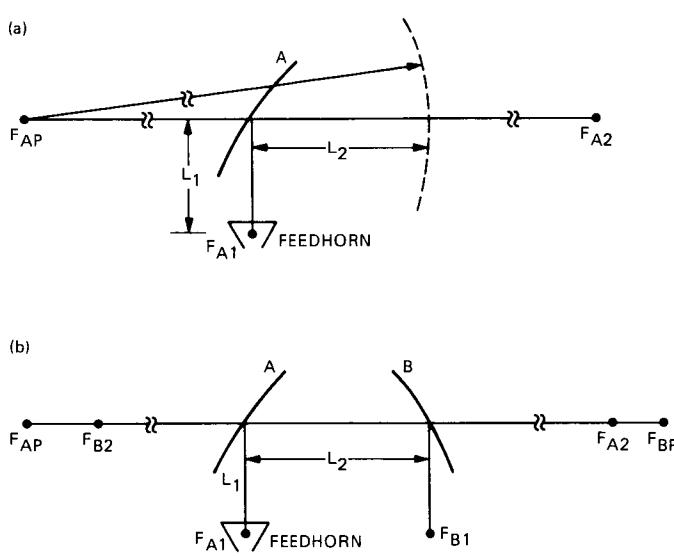
$\Downarrow$   $(S_2, S_3)$  SATISFIES MIZUSAWA  
AND IS ELIMINATED



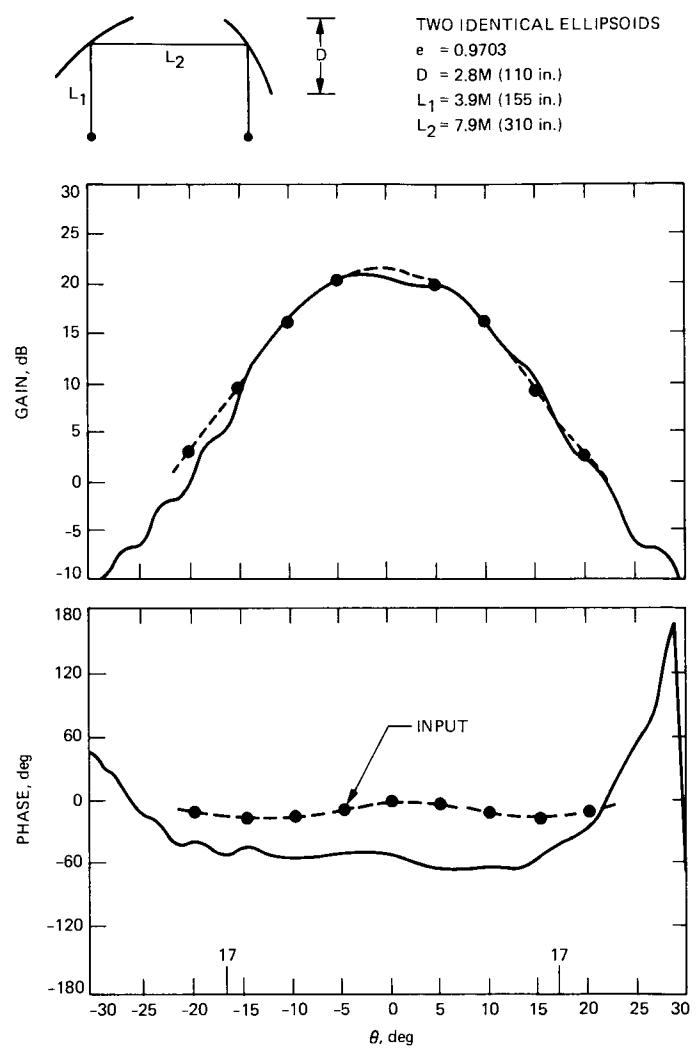
**Fig. 4. Demonstrating an extension of Mizusawa's criteria for a multiple reflector beam waveguide (ellipsoids/hyperboloids)**



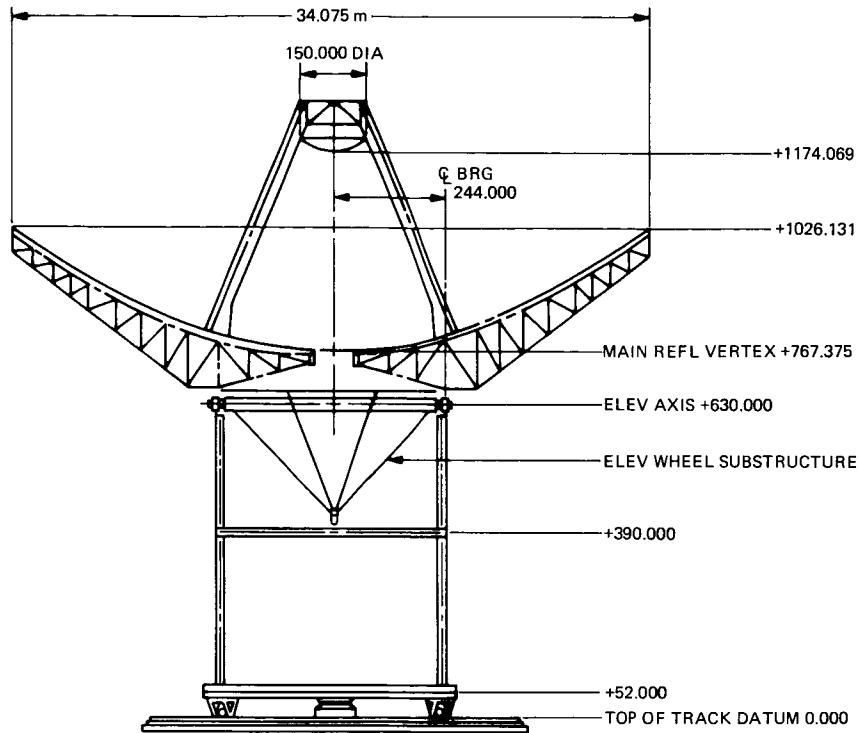
**Fig. 5. Geometrical optics field reflected from each surface of a beam-waveguide system (shown in Fig. 3[a]): (a) input pattern; (b) reflected field from  $S_1$ ; (c) reflected field from  $S_2$ ; (d) reflected field from  $S_3$ ; (e) reflected field from  $S_4$  or output pattern**



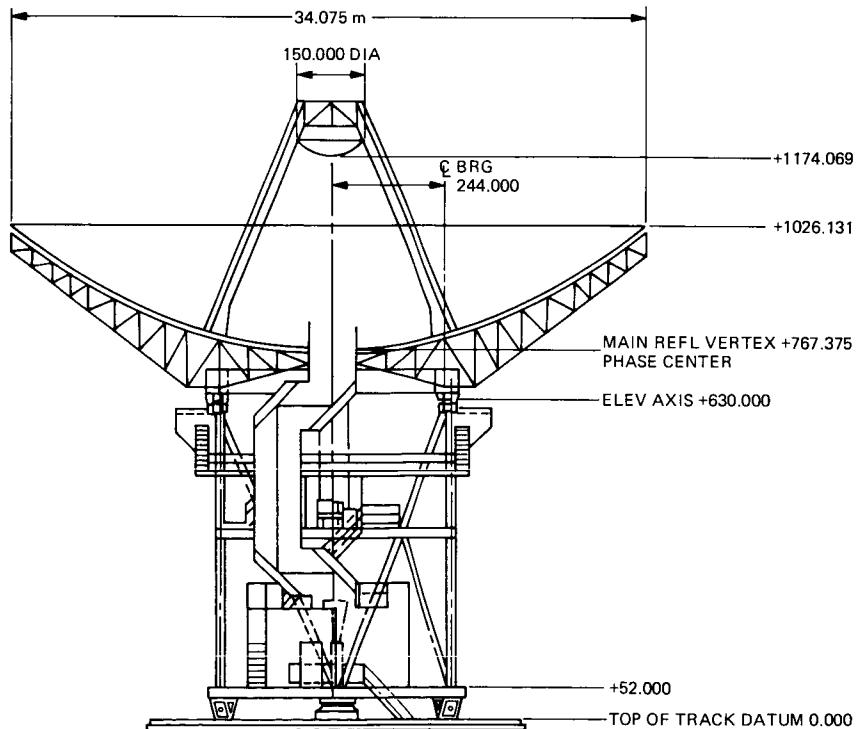
**Fig. 6. Bandpass feed imaging configuration**



**Fig. 7. Comparison between input feed pattern and feed image from a bandpass beam-waveguide system at center frequency = 2.3 GHz**



**Fig. 8. 34-m H.E. existing structure**



**Fig. 9. 34-m H.E. centerline beam waveguide**

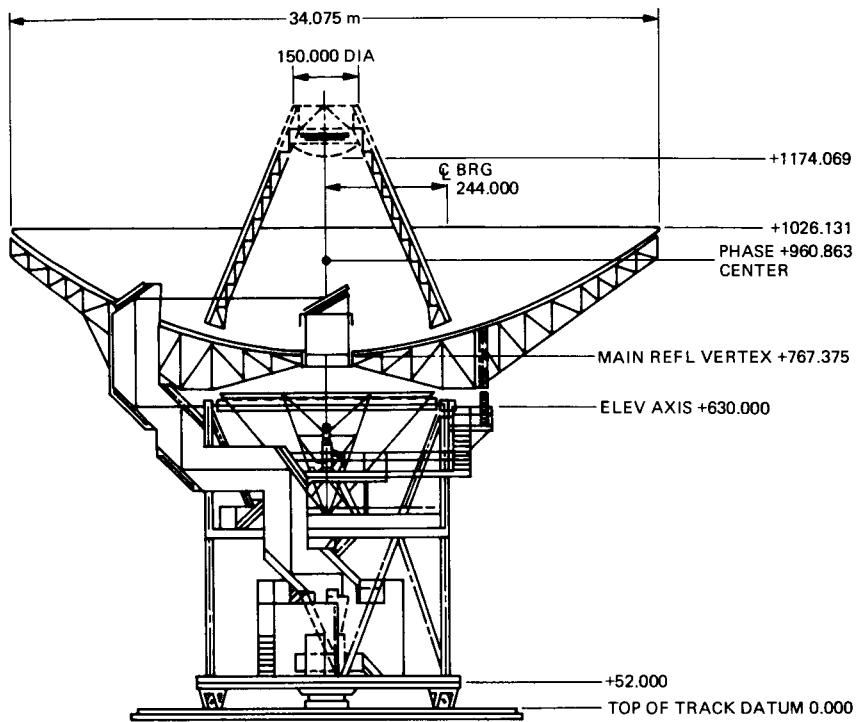


Fig. 10. 34-m H.E. bypass beam waveguide

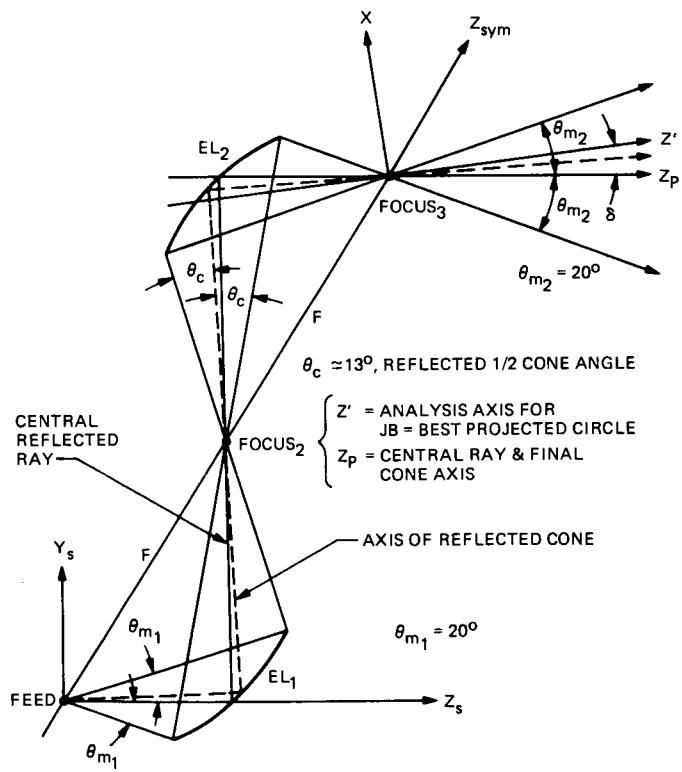
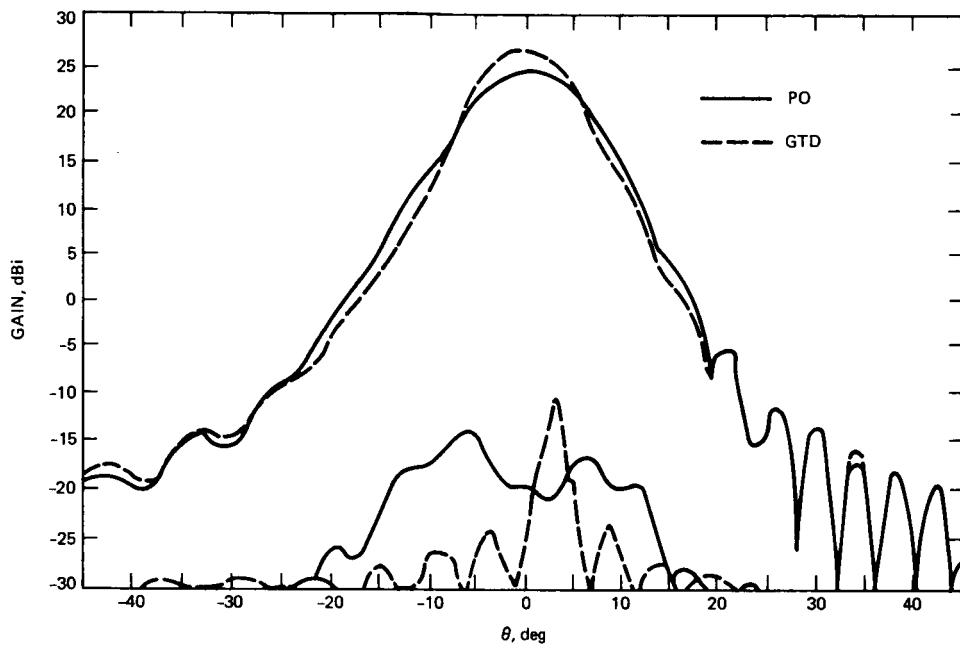
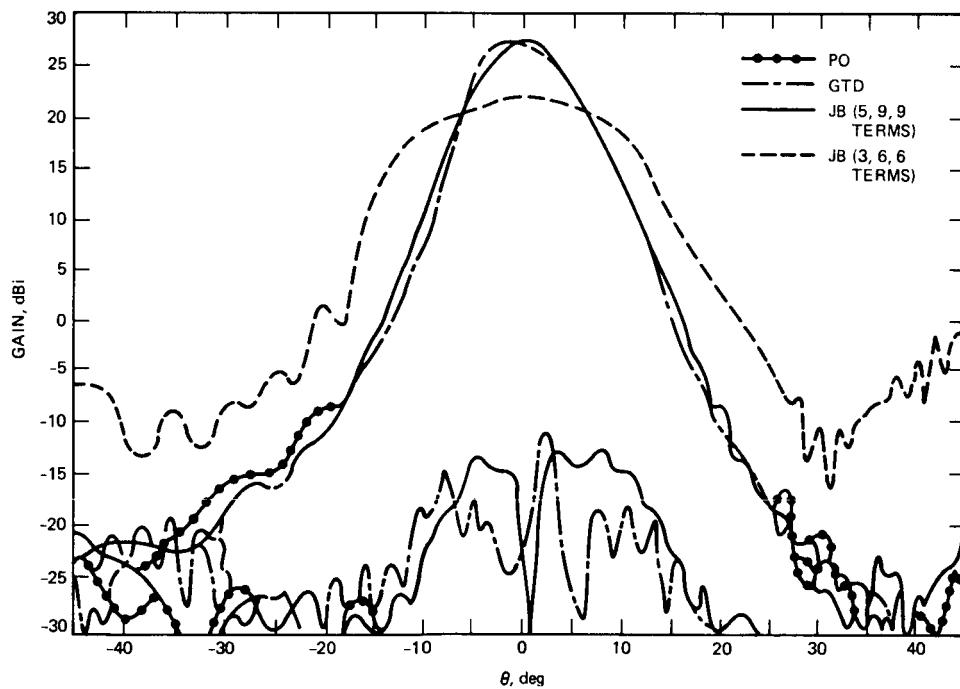


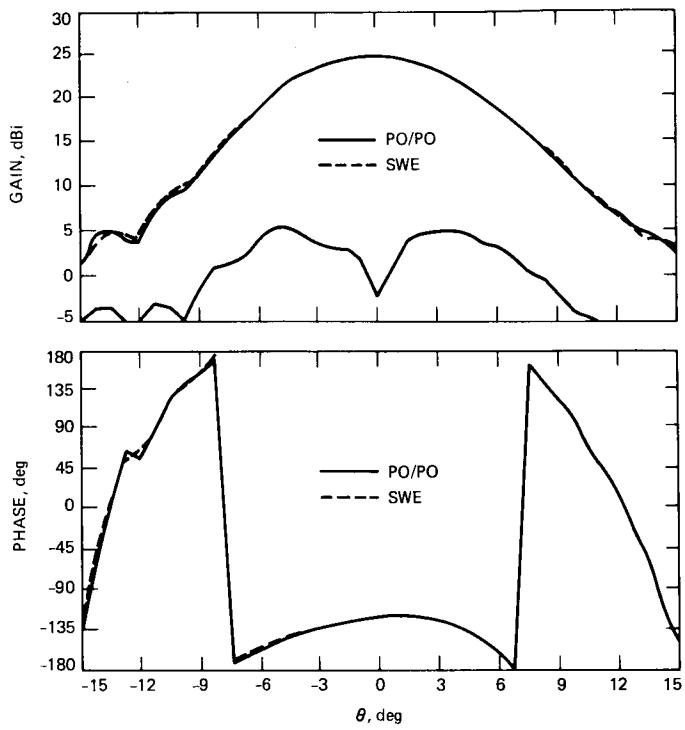
Fig. 11. Analysis geometry for ellipsoid pair ( $EL_1, EL_2$ )



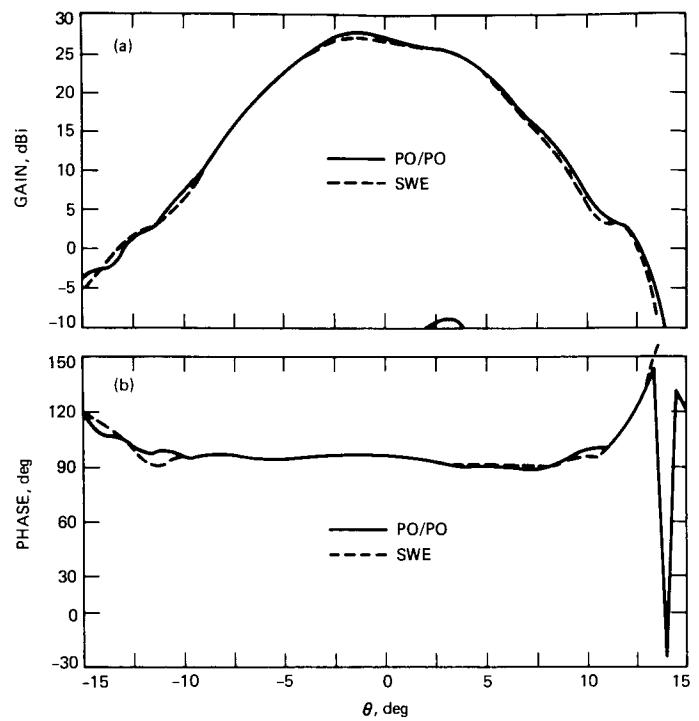
**Fig. 12. Near-field, one 2.4-m ellipse diffraction patterns (RCP, 2.3-GHz, offset plane)**



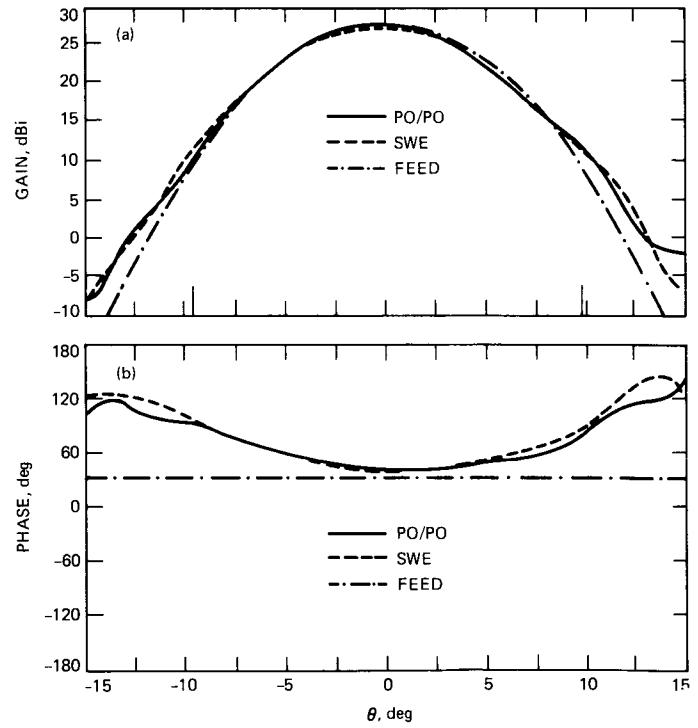
**Fig. 13. Far-field, one 2.4-m ellipse diffraction patterns (RCP, 2.3-GHz, offset plane)**



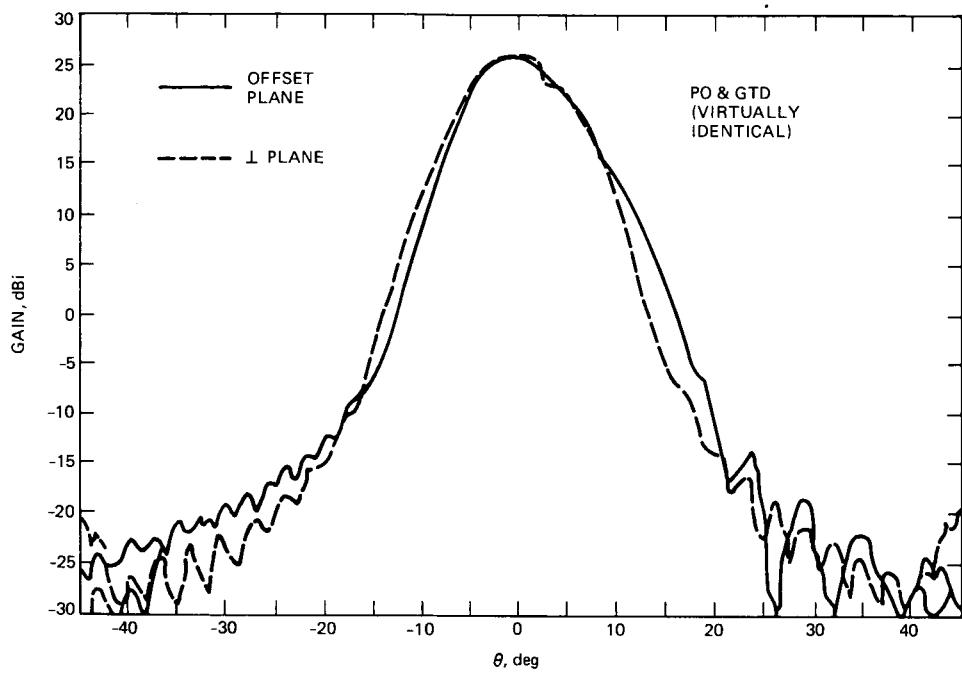
**Fig. 14. PO and SWE near-field scattering from a single 2.4-m paraboloid at 2.3 GHz: (a) amplitude; (b) phase**



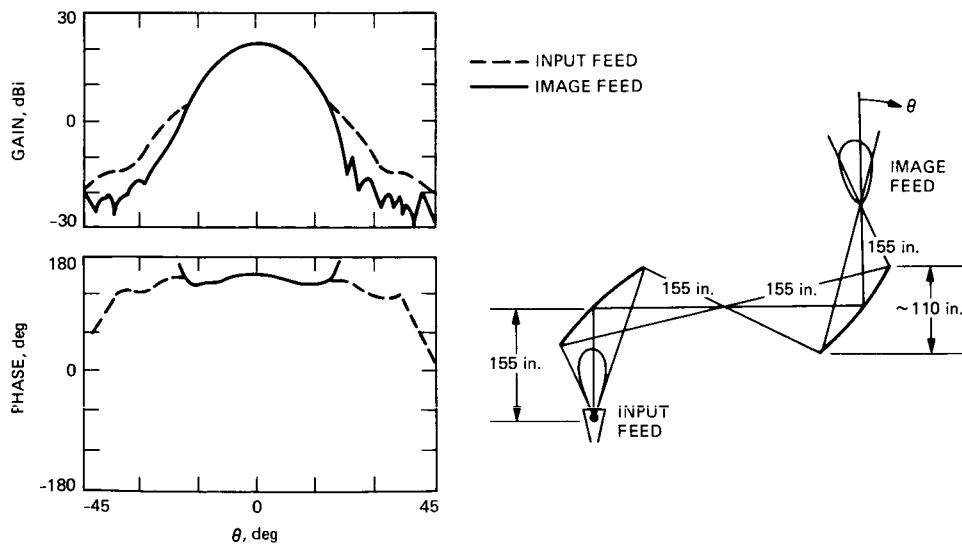
**Fig. 15. PO and SWE far-field scattering from two 2.4-m paraboloids at 2.3 GHz: (a) amplitude; (b) phase**



**Fig. 16. PO and SWE near-field scattering from two 2.4-m paraboloids at 2.3 GHz (comparison with feed pattern): (a) amplitude; (b) phase**



**Fig. 17. Near-field, one 2.4-m ellipse diffraction patterns (RCP, 8.4 GHz, offset and plane cuts)**



**Fig. 18. GTD scattering for two 2.4-m ellipses, RCP at 8.4 GHz**

# A Synchronization Technique for Optical PPM Signals

V. A. Vilnrotter, E. R. Rodemich and H. H. Tan  
Communications Systems Research Section

*A technique for maintaining synchronization between optical PPM pulses and a receiver clock by means of a delay-tracking loop is described and analyzed. The tracking loop is driven by a doubly stochastic Poisson process that contains information about the location of the desired slot boundaries. The slot boundaries are subject to slowly varying random delays that are ultimately tracked by the loop. The concept of fractional rms delay error is introduced to quantify the effects of signal and background induced shot noise on the performance of the delay-tracking loop.*

## I. Introduction

Optical pulse-position modulation (PPM) is an accepted signaling format for the transmission of information over optical channels. With this coding technique,  $L$  bits of information are encoded onto one of  $M = 2^L$  PPM words by establishing a one-to-one correspondence between the possible states of  $L$  binary digits and the location of an optical pulse among  $M$  possible slots. If the slot duration is  $\tau$  seconds, then each PPM word requires  $M\tau$  seconds for transmission. For a direct-detection system, maximum likelihood decoding consists of counting the number of photons in each slot and selecting the PPM word corresponding to the greatest count. To accomplish this, however, requires accurate synchronization between the received PPM slot boundaries and the counting boundaries defined by the receiver clock. Synchronization for binary PPM optical signals by means of "early-late" gates and decision-directed feedback has been considered previously in the literature (Refs. 1, 2). Here we consider a novel technique for synchronizing  $M$ -ary PPM symbols without the use of decision-directed feedback, thus effectively removing the dependence of the synchronization subsystem on

decoder performance. The description and analysis of this tracking system is the subject of the following sections.

## II. Tracking System Description and Analysis

The slot synchronization technique described here relies on the addition of a small "dead time" before and after each optical pulse, in order to derive an error signal proportional to the delay error that may exist between the received slot boundaries and the receiver clock. (Decoder performance is not affected if the pulse intensity at the transmitter can be increased to maintain constant average pulse count at the receiver.) Thus,  $\tau = \tau_p + 2\tau_d$ , where  $\tau_p$  is the duration of the optical pulse and  $\tau_d$  the duration of each "dead-time" interval.

The detector output can be modeled as a doubly stochastic Poisson process, represented by a sequence of randomly occurring unit impulses whose average rate at any given time depends on the total signal and background power collected by the

receiver. Consider the case where the receiver clock is synchronized with the received optical data (we do not address the acquisition problem here). Doppler effects and transmitter or receiver clock instabilities tend to introduce random delay components between the received data and the receiver clock which, if uncorrected, may eventually cause loss of synchronization. Assuming that the tracking system maintains synchronism between the received data and the receiver clock, transmitter clock instabilities and Doppler-induced rate changes can be modeled in terms of the delay process  $\Delta\tau = \Delta\tau_{cr} + \Delta\tau_D$ , whose components are transmitter clock drift and Doppler-induced delay, respectively. Typically,  $\Delta\tau$  is a slowly varying function of time that can be treated as a constant over a great many PPM words.

A block diagram of the delay-tracking loop is shown in Fig. 1. The detector output process  $x(t;\Delta\tau)$ , which contains information about the location of the transmitted data-pulses, serves as input to the tracking loop. This is a doubly stochastic Poisson process with random rate  $\lambda(t;\Delta\tau) = \lambda_s(t;\Delta\tau) + n_b$ , where  $\lambda_s(t;\Delta\tau)$  is a random rate due to the signal while  $n_b$  is a constant average count rate due to multimode background radiation. The receiver clock generates a squarewave  $g(t;\Delta\tau_n)$  of nominal period  $\tau$  and noisy delay estimate  $\widehat{\Delta\tau}_n = \widehat{\Delta\tau} + \Delta\tau_{cr}$ . The components of  $\widehat{\Delta\tau}_n$  are the loop's estimate of the received delay  $\widehat{\Delta\tau}$ , and an independent drift component  $\Delta\tau_{cr}$  due to instabilities within the receiver clock. The input to the loop filter is the product of the detector output process and clock waveform. Defining the delay error as  $\Delta\tau_\epsilon = \Delta\tau - \widehat{\Delta\tau}_n$ , this process can be represented as

$$y(t;\Delta\tau_\epsilon) = x(t;\Delta\tau) g(t;\widehat{\Delta\tau}_n) \quad (1)$$

where the notation implies that the product contains information about the delay error  $\Delta\tau_\epsilon$ . Indeed, for  $M$ -ary PPM symbols with dead-time, the time averaged mean value given  $\Delta\tau$  and  $\widehat{\Delta\tau}_n$ , clearly depends on  $\Delta\tau_\epsilon$ :

$$\begin{aligned} \bar{y}(\Delta\tau_\epsilon) &\triangleq \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T E \{x(t;\Delta\tau) g(t;\widehat{\Delta\tau}_n)\} dt \\ &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \lambda(t;\Delta\tau) g(t;\widehat{\Delta\tau}_n) dt \\ &= \frac{2n_s}{M\tau} \Delta\tau_\epsilon = \alpha \Delta\tau_\epsilon; |\Delta\tau_\epsilon| \leq \delta \end{aligned} \quad (2a)$$

$$\delta \triangleq \min \left( \tau_d, \frac{\tau}{2} - \tau_d \right) \quad (2b)$$

where  $2\delta$  is the extent of the linear range,  $E$  is the expectation operator,  $n_s$  is the Poisson rate due to signal when a pulse is present, and  $\alpha = 2n_s/M\tau$ . Note that  $\bar{y}(\Delta\tau_\epsilon)$  does not depend on the background induced count rate since the time average of  $n_b g(t;\widehat{\Delta\tau}_n)$  over an integral number of periods is zero. Figure 2 illustrates the behavior of  $\bar{y}(\Delta\tau_\epsilon)$  as a function of  $\Delta\tau_\epsilon$  over the interval  $|\Delta\tau_\epsilon| \leq \tau/2$ , assuming that  $\tau_d < \tau/4$ .

It is convenient to decompose  $y(t;\Delta\tau_\epsilon)$  as

$$y(t;\Delta\tau_\epsilon) = \bar{y}(\Delta\tau_\epsilon) + n(t;\Delta\tau_\epsilon) \quad (3)$$

where  $n(t;\Delta\tau_\epsilon) \triangleq y(t;\Delta\tau_\epsilon) - \bar{y}(\Delta\tau_\epsilon)$  is viewed as an additive noise process, containing the conditioning parameter  $\Delta\tau_\epsilon$ . The power spectral density of this noise process conditioned on  $\Delta\tau_\epsilon$ ,  $S_n(\omega;\Delta\tau_\epsilon)$ , is derived in Appendix A. There it is shown that for  $\omega = 0$  the noise power spectral density does not depend on  $\Delta\tau_\epsilon$ , but takes the value  $(\tilde{n}_s + n_b)$ , where

$$\tilde{n}_s \triangleq n_s \tau_p / M\tau$$

is the average Poisson count rate generated by the signal. Thus, in a suitably narrow band of frequencies around the origin, the power spectral density of the equivalent noise process may be approximated as

$$S_n(\omega) \approx \tilde{n}_s + n_b; |\omega| \ll \frac{2\pi}{M\tau} \quad (4)$$

If  $\Delta\tau_\epsilon$  varies slowly enough to be considered constant over a great many PPM symbols, then  $\bar{y}(\Delta\tau_\epsilon)$  and  $S_n(\omega;\Delta\tau_\epsilon)$  may be approximated by replacing the infinite limits in their defining equations by finite values that encompass regions of constant  $\Delta\tau_\epsilon$ . In that case, we conclude that for frequencies near the origin, the process  $y(t;\Delta\tau_\epsilon)$  behaves as though it were composed of a slowly varying term proportional to the delay error; plus an independent additive white noise process  $n(t)$  characterized by the spectral level  $(\tilde{n}_s + n_b)$ , as defined in equation (4).

Assuming that  $|\Delta\tau_\epsilon| \ll \delta$ , the operation of the tracking loop can be described by means of the linear loop model shown in Fig. 3 (Ref. 3). Using the differential operator  $p \triangleq d/dt$ , loop operation can be expressed in terms of the equation

$$\begin{aligned} \Delta\hat{\tau} &= \Delta\tau_c + \Delta\tau_D - \Delta\tau_\epsilon = \frac{G}{P} F(p) [\alpha\Delta\tau_\epsilon + n] \\ \Delta\tau_c &\triangleq \Delta\tau_{ct} - \Delta\tau_{cr} \end{aligned} \quad (5)$$

where  $G$  is the gain of the voltage-controlled clock and  $n = n(t;\Delta\tau_\epsilon)$  as in Eq. (3) (the time index has been deleted for the sake of notational simplicity). Solving for  $\Delta\tau_\epsilon$  yields

$$\Delta\tau_\epsilon = [1 - H(p)] (\Delta\tau_c + \Delta\tau_D) - H(p) \frac{n}{\alpha} \quad (6a)$$

$$H(p) \triangleq \alpha GF(p)/(p + \alpha GF(p)) \quad (6b)$$

The tracking system should be designed to track  $\Delta\tau_c$  and  $\Delta\tau_D$  accurately, in which case their contributions to  $\Delta\tau_\epsilon$  can be ignored. Under this condition the steady-state delay-error is due entirely to the equivalent noise process  $n(t)$ . Assuming that  $|H(j\omega)|^2$  takes on significant values only in an interval  $|\omega| \ll 2\pi/M\tau$ , the variance of the delay-error becomes

$$\begin{aligned} \sigma_\epsilon^2 \triangleq \text{var}(\Delta\tau_\epsilon) &= \frac{1}{2\pi\alpha^2} \int_{-\infty}^{\infty} |H(j\omega)|^2 S_n(\omega) d\omega \\ &\simeq \frac{2B_L}{\alpha^2} S_n(0) \end{aligned} \quad (7)$$

where  $B_L$  is the one-sided loop bandwidth defined by

$$2B_L = \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(j\omega)|^2 d\omega \quad (8)$$

Substituting for  $\alpha$  and  $S_n(0)$  in equation (7) yields

$$\sigma_\epsilon^2 = \frac{B_L M^2 \tau^2}{2 n_s^2} (\tilde{n}_s + n_b) = \frac{B_L \tau_p^2}{2 \tilde{n}_s^2} (\tilde{n}_s + n_b) \quad (9)$$

From equation (9) it is clear that the fractional rms delay error  $(\sigma_\epsilon/\tau_p)$  depends only on  $B_L$ ,  $\tilde{n}_s$  and  $n_b$ , and is independent of  $\tau_p$  for any fixed value of  $\tilde{n}_s$ :

$$\left(\frac{\sigma_\epsilon}{\tau_p}\right) = \sqrt{\frac{B_L}{2\tilde{n}_s}} \left(1 + \frac{n_b}{\tilde{n}_s}\right) \quad (10)$$

Note that even in the absence of background radiation  $(\sigma_\epsilon/\tau_p)$  does not become arbitrarily small, but approaches its quantum-limited value  $\sqrt{B_L/2\tilde{n}_s}$ .

It is apparent from equations (6b) and (8) that  $B_L$  depends only on the loop filter transfer function  $F(p)$  and the product  $\alpha G$ . Since  $G$  is independent of the other system parameters, constant loop bandwidth can be maintained by adjusting  $G$  so as to keep  $\alpha G$  constant. Thus,  $B_L$  can be held fixed even as  $\tilde{n}_s = \alpha\tau_p/2$  varies. The behavior of  $(\sigma_\epsilon/\tau_p)$  as a function of  $\tilde{n}_s$  for various loop bandwidths and background rates is shown in Fig. 4.

Other meaningful fractional rms delay errors may also be defined in terms of  $(\sigma_\epsilon/\tau_p)$ , namely

$$\left(\frac{\sigma_\epsilon}{\tau_d}\right) = \left(\frac{\tau}{\tau_d} - 2\right) \left(\frac{\sigma_\epsilon}{\tau_p}\right); \quad 0 < \tau_d \leq \frac{\tau}{2} \quad (11a)$$

$$\left(\frac{\sigma_\epsilon}{\tau}\right) = \left(1 - \frac{2\tau_d}{\tau}\right) \left(\frac{\sigma_\epsilon}{\tau_p}\right); \quad 0 < \tau_d \leq \frac{\tau}{2} \quad (11b)$$

and using Eq. (2b),

$$\left(\frac{\sigma_\epsilon}{\delta}\right) = \begin{cases} \left(\frac{\sigma_\epsilon}{\tau_d}\right); & 0 < \tau_d \leq \frac{\tau}{4} \\ 2\left(\frac{\sigma_\epsilon}{\tau_p}\right); & \frac{\tau}{4} < \tau_d \leq \frac{\tau}{2} \end{cases} \quad (11c)$$

It should be emphasized that the above equations are meaningful only if the inequality  $(\sigma_\epsilon/\delta) \ll 1$  is satisfied. Assuming this to be the case, Eqs. (11a) and (11c) show that  $(\sigma_\epsilon/\delta)$  is a monotone decreasing function of  $\tau_d$  in the interval  $0 < \tau_d \leq \tau/4$ , while maintaining a constant minimum value of  $2(\sigma_\epsilon/\tau_p)$  over the interval  $\tau/4 < \tau_d < \tau/2$ .

In typical applications, the loop bandwidth  $B_L$  is chosen large enough to exceed the significant spectral components of  $\Delta\tau_c$  and  $\Delta\tau_D$ , thus allowing the loop to track these quantities accurately. The signal-set dimension  $M$  and slot-width  $\tau$  are generally determined on the basis of communications requirements, while  $n_b$  depends on the background environment and on the particular set of spatial and frequency predetection filters employed. In order to achieve the desired decoder error probability, some minimum average signal pulse count  $n_s \tau_p$  must be maintained, even in the absence of synchronization errors. Under the condition  $(\sigma_\epsilon/\delta) \ll 1$ , random delay errors should have little effect on decoder performance if the decoder counts photons over the entire synchronous  $\tau$ -second time interval. Thus, a value of  $(\sigma_\epsilon/\delta)$  can generally be found to guarantee negligible decoder performance degradation while simultaneously assuring linear loop operation: let this value be denoted by  $\gamma$ . Again using Eqs. (11a) and (11c), a reasonable value for  $\tau_d$  can be determined by calculating  $(\sigma_\epsilon/\tau_p)$  from the given system parameters, and comparing it to  $\gamma/2$ . If  $(\sigma_\epsilon/\tau_p) \leq \gamma/2$ , then let

$$\tau_d = \tau \left[ 2 + \left( \gamma / \left( \frac{\sigma_\epsilon}{\tau_p} \right) \right) \right]^{-1} \quad (12)$$

However, if  $(\sigma_e/\tau_p) > \gamma/2$ , then the desired value of  $(\sigma_e/\delta)$  cannot be realized with the initial set of system parameters, and hence at least one of the original parameters must be altered to accommodate tracking-loop requirements. Generally, the tracking loop will perform well, with dead-times substantially less than  $\tau/4$  seconds, when operating in high data-rate narrow band communication systems where a suitably small value of  $(\sigma_e/\tau_p)$  can be achieved.

### III. Concluding Remarks

A technique has been developed for maintaining synchronization between a PPM-modulated optical data-stream and a receiver clock. This technique relies on the addition of a small "dead time" interval before and after each possible pulse location to derive an electrical signal proportional to the delay error between the received slot boundaries and the receiver clock. It was argued that in applications where only spectral components near the origin are of interest, and where the delay error remains essentially constant over a great many

PPM symbols, the driving process could be approximated by a slowly varying term proportional to the delay error, plus an effective additive noise process due to the inherent randomness of the detected optical fields. Based on the above approximations, the variance of the delay error due to both signal and background radiation was derived, assuming a linear loop that effectively tracked out the slowly varying delay error components. We observed that since linear loop operation typically implied rms delay errors on the order of a small fraction of the linear region, decoder performance should not be significantly affected if the decoder counts photons over the entire  $\tau$ -second slot time. It was shown that in applications where this is true, a simple algorithm can often be employed to find "good" design values for the dead-time interval in terms of the given system parameters, and a quantity the designer selects to simultaneously ensure linear loop operation and acceptable decoder performance. However, a precise determination of the dead-time interval (taking into account decoder performance degradation and possible non-linear effects within the loop) is beyond the scope and interest of this article.

### References

1. Gagliardi, R. M., "Synchronization Using Pulse Edge Tracking in Optical Pulse-Position Modulated Communication Systems," *IEEE Transactions on Communications*, Oct. 1974, Vol. *Com-22*, No. 10, pp. 1693-1702.
2. Gagliardi, R. M. and Karp, S., *Optical Communications*, Wiley, N.Y., 1976.
3. Lindsey, W. C., *Synchronization Systems in Communications and Control*, Prentice Hall, N.J., 1972, Chapt. 4.

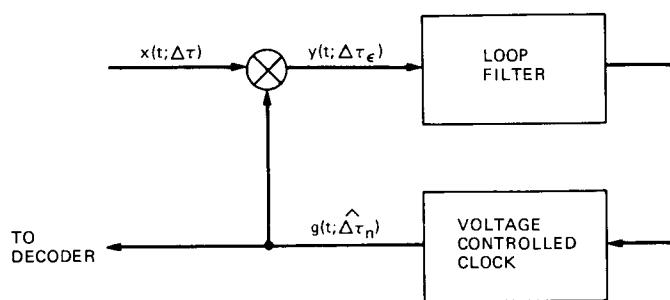


Fig. 1. Tracking loop block diagram

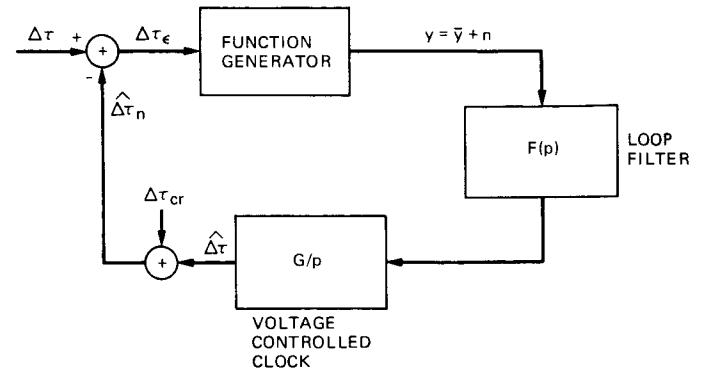


Fig. 3. Linear loop model

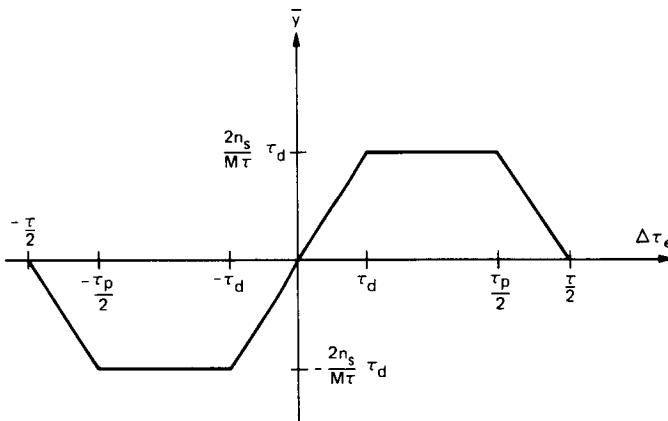


Fig. 2. Time averaged mean value as a function of delay error ( $\tau_d < \tau/4$ )

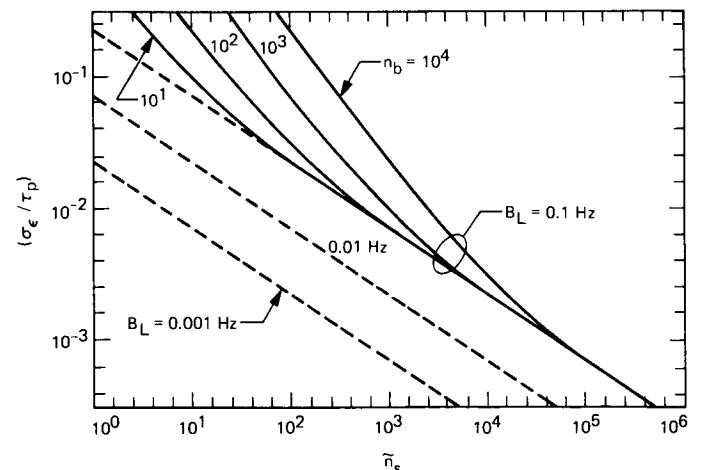


Fig. 4. Fractional rms delay error  $(\sigma_\epsilon / \tau_p)$  as a function of average signal count for various background rates and loop bandwidths

## Appendix A

The power spectral density of  $n(t; \Delta\tau_\epsilon)$  can be derived using the results of Appendix B. Assume that  $\Delta\tau_\epsilon$  is held fixed. Since  $n(t; \Delta\tau_\epsilon) = y(t; \Delta\tau_\epsilon) - \bar{y}(\Delta\tau_\epsilon)$ , it follows that

$$S_n(\omega; \Delta\tau_\epsilon) = S_y(\omega; \Delta\tau_\epsilon) - S_{\bar{y}}(\omega; \Delta\tau_\epsilon) \quad (\text{A-1})$$

where

$$S_{\bar{y}}(\omega; \Delta\tau_\epsilon) = \left( \frac{2n_s \Delta\tau_\epsilon}{M\tau} \right)^2 2\pi\delta(\omega)$$

Consider the case where the detector observes signal plus multimode background noise fields. Thus,  $\lambda(t; \Delta\tau) = \lambda_s(t; \Delta\tau) + n_b$  where  $\lambda_s(t; \Delta\tau)$  is the random rate function associated with the signal while  $n_b$  is a constant average count rate contributed by the background fields. The square-wave generated by the receiver clock is denoted by  $g(t; \Delta\tau_n)$  as before. For M-ary PPM with dead-time the product  $\lambda(t; \Delta\tau) g(t; \Delta\tau_n)$  can be modeled over the interval  $[-T, T]$  as

$$\begin{aligned} \lambda(t; \Delta\tau) g(t; \Delta\tau_n) &= \sum_{i=-K}^K p_s \left( t - \left( iM\tau + \ell_i \tau - \frac{M\tau}{2} \right) \right) \\ &+ \sum_{i=-K}^K \sum_{k=0}^{M-1} p_b \left( t - \left( iM\tau + k\tau - \frac{M\tau}{2} \right) \right) \end{aligned} \quad (\text{A-2})$$

where  $K = (T/M\tau) \sim 1/2$ ,  $\tau$  is the slot duration, and  $M\tau$  the duration of a PPM word. Elements of the set  $\{\ell_i\}$  are independent random integers taking on values from 0 to  $(M - 1)$  with equal probability. Given  $\Delta\tau_\epsilon$ ,  $p_s(t)$  and  $p_b(t)$  are defined as

$$p_s(t) = \begin{cases} 0; 0 \leq t < \tau_d, \tau - \tau_d \leq t < \tau \\ n_s; \tau_d \leq t < \frac{\tau}{2} + \Delta\tau_\epsilon \\ -n_s; \frac{\tau}{2} + \Delta\tau_\epsilon \leq t < \tau - \tau_d \end{cases} \quad (\text{A-3a})$$

$$p_b(t) = \begin{cases} n_b; \Delta\tau_\epsilon \leq t < \frac{\tau}{2} + \Delta\tau_\epsilon \\ -n_b; 0 \leq t < \Delta\tau_\epsilon, \frac{\tau}{2} + \Delta\tau_\epsilon \leq t < \tau \end{cases} \quad (\text{A-3b})$$

With  $P_s(\omega) \triangleq F\{p_s(t)\}$  and  $P_b(\omega) \triangleq F\{p_b(t)\}$ , the Fourier transform of  $\lambda(t; \Delta\tau) g(t; \Delta\tau_n)$  restricted to the interval  $[-T, T]$ , becomes

$$\begin{aligned} F_T\{\lambda(t; \Delta\tau) g(t; \Delta\tau_n)\} &= \left\{ P_s(\omega) \sum_{i=-K}^K e^{-j\omega(iM\tau + \ell_i\tau)} \right. \\ &\quad \left. + MP_b(\omega) \gamma(\omega) \right\} e^{j\omega M\tau/2} \end{aligned} \quad (\text{A-4})$$

where

$$\begin{aligned} \gamma(\omega) &= \sum_{i=-K}^K e^{-j\omega iM\tau} E\left(e^{j\omega \ell_i \tau}\right) \\ &= e^{-(1/2)j\omega(M-1)\tau} \left[ \frac{\sin\left(\left(K + \frac{1}{2}\right)\omega M\tau\right)}{M \sin\left(\frac{1}{2}\omega\tau\right)} \right] \\ P_s(\omega) &= \frac{n_s}{j\omega} \left[ e^{-j\omega\tau_d} + e^{-j\omega(\tau-\tau_d)} - 2e^{-j\omega(\tau/2+\Delta\tau_\epsilon)} \right] \end{aligned}$$

and

$$P_b(\omega) = \frac{n_b}{j\omega} \left[ 2e^{-j\omega\Delta\tau_\epsilon} + e^{-j\omega\tau} - 2e^{-j\omega(\tau/2+\Delta\tau_\epsilon)} - 1 \right]$$

Recognizing that

$$\sum_{i=-K}^K \sum_{n=-K}^K e^{-j\omega(i-n)M\tau} E\left\{ e^{-j\omega(\ell_i - \ell_n)\tau} \right\}$$

$$= |\gamma(\omega)|^2 + (2K+1) \left[ 1 - \left( \frac{\sin\left(\frac{1}{2}\omega M\tau\right)}{M \sin\left(\frac{1}{2}\omega\tau\right)} \right)^2 \right]$$

and making use of Eq. (A-4), it follows that

$$\frac{1}{2T} E \left\{ |F_T \{\lambda g\}|^2 \right\} = \frac{1}{2T} \left\{ |P_s(\omega) + MP_b(\omega)|^2 |\gamma(\omega)|^2 + |P_s(\omega)|^2 (2K+1) \left[ 1 - \left( \frac{\sin \left( \frac{1}{2} \omega M\tau \right)}{M \sin \left( \frac{1}{2} \omega \tau \right)} \right)^2 \right] \right\} \quad (\text{A-5})$$

Note that

$$\lim_{K \rightarrow \infty} \frac{|\gamma(\omega)|^2}{(2K+1)} = \frac{2\pi}{M\tau} \sum_{m=-\infty}^{\infty} \delta \left( \omega - \frac{2\pi m}{\tau} \right)$$

To show this we first observe that  $\gamma(\omega)$  is a periodic function of  $\omega$  with period  $2\pi/\tau$ . Over any interval  $I$  that does not contain the points  $\omega_m = 2\pi m/\tau$ , the denominator

$$M \sin(\omega\tau/2)$$

is bounded away from zero, hence  $|\gamma(\omega)|^2$  is bounded, and

$$\lim_{K \rightarrow \infty} \frac{|\gamma(\omega)|^2}{(2K+1)} = 0; \quad \omega_m \notin I$$

Thus, the integral of  $|\gamma(\omega)|^2/(2K+1)$  vanishes everywhere except within arbitrarily small neighborhoods around the points  $\omega_m$  as  $K$  grows without bound. However, integration over any complete cycle yields

$$\begin{aligned} \int_{2\pi(m-1/2)/\tau}^{2\pi(m+1/2)/\tau} d\omega |\gamma(\omega)|^2 / (2K+1) &= \frac{1}{M^2 (2K+1)} \sum_{i=-K}^K \sum_{\ell=0}^{M-1} \frac{2\pi}{\tau} \\ &= \frac{2\pi}{M\tau} \end{aligned}$$

independent of  $K$ . Therefore, this is the value of the integral over arbitrarily small neighborhoods around each point  $\omega_m$  as  $K$  approaches infinity.

We also make the useful observations that  $P_b(0) = 0$ ,  $|P_s(0)|^2 = 4 n_s^2 \Delta\tau_e^2$ , and

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T E \{|g(t; \widehat{\Delta\tau}_n)|^2 \lambda(t; \Delta\tau) dt \\ = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \lambda(t; \Delta\tau) dt \\ = \widetilde{n}_s + n_b \end{aligned}$$

where

$$\widetilde{n}_s = n_s \tau_p / M\tau \quad \text{and} \quad |g(t; \widehat{\Delta\tau}_n)|^2 = 1$$

Making use of the fact that  $2T = (2K+1)M\tau$ , substituting the above terms into equation (B-7) and using Eq. (A-1) finally yields

$$\begin{aligned} S_n(\omega; \Delta\tau_e) &= \widetilde{n}_s + n_b + \frac{|P_s(\omega)|^2}{M\tau} \left[ 1 - \left( \frac{\sin \left( \frac{1}{2} \omega M\tau \right)}{M \sin \left( \frac{1}{2} \omega \tau \right)} \right)^2 \right] \\ &\quad + |P_s(\omega) + MP_b(\omega)|^2 \frac{2\pi}{M^2 \tau^2} \sum_{\substack{m=-\infty \\ m \neq 0}}^{\infty} \delta \left( \omega - \frac{2\pi m}{\tau} \right) \end{aligned} \quad (\text{A-6})$$

Thus, the spectral density of the equivalent noise process consists of continuous and impulsive terms. The impulses occur at integer multiples of the radian frequency  $2\pi/\tau$ ; however, there is no impulse at the origin since the contribution of  $\bar{y}$  was subtracted out. The value of the continuous part at the origin is

$$S_n(0; \Delta\tau_e) = \widetilde{n}_s + n_b \quad (\text{A-7})$$

independent of  $\Delta\tau_e$ . Therefore, this must be the value of the unconditioned spectral density as well:

$$S_n(0) = E_{\Delta\tau_e} \{S_n(0; \Delta\tau_e)\} = \widetilde{n}_s + n_b \quad (\text{A-8})$$

Since for any  $|\Delta\tau_e| \leq \delta$ , the continuous part of Eq. (A-6) begins to deviate significantly from its zero-frequency value only as  $\omega$  approaches values on the order of  $2\pi/M\tau$ , it can be argued that in a suitably narrow band of frequencies around the origin (namely,  $|\omega| \ll 2\pi/M\tau$ ) the spectral density of the equivalent noise process may be approximated by its zero-frequency value.

## Appendix B

The power spectral density of the product of a doubly stochastic Poisson process and a deterministic function is derived. Let

$$x(t) = \sum_{\ell} \delta(t - t_{\ell}) \quad (\text{B-1})$$

be a sample function of a doubly stochastic Poisson process with random rate  $\lambda(t)$ . Let  $g(t)$  be a deterministic function and form the product  $y(t) = g(t) x(t)$ . Define the power spectral density of  $y(t)$ ,  $S_y(\omega)$ , as in Ref. 2 (Chapter 4):

$$S_y(\omega) = \lim_{T \rightarrow \infty} \frac{1}{2T} E |Y_T(\omega)|^2 \quad (\text{B-2})$$

where

$$Y_T(\omega) \triangleq \int_{-T}^T e^{-j\omega t} y(t) dt = F_T \{y(t)\}$$

is the Fourier transform of  $g(t)$  restricted to the interval  $[-T, T]$ . Using Eq. (B-1) yields

$$Y_T(\omega) = \sum_{\ell} g(t_{\ell}) e^{-j\omega t_{\ell}}; \quad t \in [-T, T] \quad (\text{B-3a})$$

$$|Y_T(\omega)|^2 = \sum_{\ell} \sum_m g(t_{\ell}) g^*(t_m) e^{j\omega(t_m - t_{\ell})}; \quad t \in [-T, T] \quad (\text{B-3b})$$

First assume that  $\lambda(t)$  is known. For a given  $\lambda(t)$  and conditioned on  $k$  arrivals in  $[-T, T]$ , the joint probability density of the unordered arrival times  $t_k = \{t_1, t_2, \dots, t_k\}$  can be expressed as

$$p_{t_k}(t_1, t_2, \dots, t_k | k) = \prod_{i=1}^k \left( \frac{\lambda(t_i)}{m_T} \right); \quad t \in [-T, T] \quad (\text{B-4})$$

where

$$m_T \triangleq \int_{-T}^T \lambda(t) dt$$

Thus,

$$\begin{aligned} E_{t_k} \{|Y_T(\omega)|^2 | k\} &= \sum_{\ell} \sum_m \int_{-T}^T \cdots \int_{-T}^T g(t_{\ell}) g^*(t_m) \\ &\times e^{j\omega(t_m - t_{\ell})} \prod_{i=1}^k \left( \frac{\lambda(t_i)}{m_T} \right) dt_1 dt_2 \dots dt_k \\ &= k \int_{-T}^T |g(t)|^2 \frac{\lambda(t)}{m_T} dt \\ &+ \frac{(k^2 - k)}{m_T^2} \int_{-T}^T dt_{\ell} g(t_{\ell}) \lambda(t_{\ell}) e^{-j\omega t_{\ell}} \\ &\times \int_{-T}^T dt_m g^*(t_m) \lambda(t_m) e^{j\omega t_m} \end{aligned} \quad (\text{B-5})$$

The first term accounts for those  $k$  elements for which  $\ell = m$ , while the second term is due to the remaining  $(k^2 - k)$  elements of the double sum. Since for a Poisson process with rate  $m_T$ ,  $E(k) = m_T$  and  $E(k^2 - k) = m_T^2$ , it follows from Eq. (B-5) that

$$\begin{aligned} E_k \left\{ E_{t_k} \{|Y_T(\omega)|^2 | k\} \right\} &= \int_{-T}^T |g(t)|^2 \lambda(t) dt \\ &+ |F_T \{g(t) \lambda(t)\}|^2 \end{aligned} \quad (\text{B-6})$$

In general  $\lambda(t)$  is itself a sample function of a random process; hence an additional averaging is required over the sample space of  $\lambda(t)$ . Thus, for doubly stochastic Poisson processes, the power spectrum is given by the expression

$$\begin{aligned} S_y(\omega) &= \lim_{T \rightarrow \infty} \frac{1}{2T} E \left\{ \int_{-T}^T |g(t)|^2 \lambda(t) dt \right. \\ &\left. + |F_T \{g(t) \lambda(t)\}|^2 \right\} \end{aligned} \quad (\text{B-7})$$

P-A

# Received Optical Power Calculations for Optical Communications Link Performance Analysis

W. K. Marshall and B. D. Burk  
 Communications Systems Research Section

*The factors affecting optical communication link performance differ substantially from those at microwave frequencies, due to the drastically differing technologies, modulation formats, and effects of quantum noise in optical communications. In addition, detailed design control table calculations for optical systems are less well developed than corresponding microwave system techniques, reflecting the relatively less mature state of development of optical communications. This article describes detailed calculations of received optical signal and background power in optical communications systems, with emphasis on analytic models for accurately predicting transmitter and receiver system losses.*

## I. Introduction

In a recent article (Ref. 1), a "simple" method for optical communications link analysis was described. That method provides a means for making quite accurate first-order predictions of optical communication link performance, without the need for involved calculation. In contrast, this article describes the detailed calculations that are necessary to more accurately predict the received optical signal and background power via the range equation. Direct-detection and pulse-position modulation (PPM) are assumed, although many of the results are independent of this assumption.

In this article, signal power is tracked from the output of the transmitter laser to the face of the detector in the receiver. The received optical background power is calculated from the background radiance. Detector sensitivity and required signal calculations are not considered here, but will be addressed in a future article.

Although some empirical factors in the calculation remain, the received power is determined here primarily from analytic models of the system components. The calculations shown here are substantially more detailed than those in earlier reports (Refs. 1-3). Future work is expected to further improve the accuracy of the calculations, and, as experience is gained with optical communications system design, to address additional issues such as design uncertainties and statistical design models (Ref. 4).

## II. Basic Parameters

In calculating link performance, we start with a set of basic component values and system parameters, which are assumed to be known and fixed in value. (As the accuracy of the link performance model is improved, some "basic" component values can, in fact, be broken down and computed from other factors. For example, with specific laser hardware,

output power can be predicted as a function of the operating conditions.)

A block diagram of an optical communications transceiver is shown in Fig. 1. A system is typically composed of two such transceivers, similar in design, though often with different component values (e.g., telescope sizes, laser powers, etc.). Additional components which are necessary for a working system but which do not directly affect communications performance (for example, the spatial acquisition system) are not shown.

The component values required to determine link performance are as follows:

- (1) Laser average power output,  $P_\ell$  (Watts)
- (2) Laser wavelength,  $\lambda$  (meters)
- (3) Transmit telescope aperture diameter,  $D_t$
- (4) Transmit telescope obscuration ratio,  $\gamma_t$
- (5) Transmitter optics throughput,  $\eta_t$
- (6) Transmitter pointing bias error,  $\epsilon_t$ , and jitter,  $\sigma_t$
- (7) Receiver telescope aperture diameter,  $D_r$
- (8) Receiver telescope obscuration ratio,  $\gamma_r$
- (9) Receiver optics throughput,  $\eta_r$
- (10) Narrow band filter transmission,  $\eta_\lambda$
- (11) Receiver pointing bias error,  $\epsilon_r$ , and jitter,  $\sigma_r$
- (12) Narrow band filter spectral bandwidth,  $\Delta\lambda$  (angstroms)
- (13) Detector diametrical field of view,  $\theta$

The transmitter and receiver optics throughput, items (5) and (9) above, can in fact be calculated from the reflection and transmission coefficients of the components in the optical system.<sup>1</sup> Table 1 gives a list of sample component values which will be used below in calculating a sample link design control table.

Besides hardware characteristics, system operational parameters are needed to determine link performance. These factors are as follows:

- (14) Data rate,  $R$  (bits/second)
- (15) PPM alphabet size,  $M$

<sup>1</sup> See, for example, Lambert, S. G., et al., "Design and Analysis Study of a Spacecraft Optical Transceiver Package," Appendix C, McDonnell Douglas Astronautics Co. (final report under JPL contract No. 957061), JPL document No. 9950-1240, Jet Propulsion Laboratory, Pasadena, Calif., Aug. 1985.

- (16) PPM slot time,  $\tau_s$  (seconds)
- (17) Link range,  $L$  (meters)
- (18) Atmospheric transmission loss factor,  $\ell_a$
- (19) Background radiance,  $B$  (Watts/meter<sup>2</sup>/steradian/angstrom)

Both the data rate, item (14), and the PPM slot time, item (16), must be specified since there may be a "dead time" between each PPM word during which the laser is always "off." This is typical in systems using Nd:YAG or other crystalline lasers. The data rate,  $R$ , slot time,  $\tau_s$ , and dead time,  $\tau_d$ , are related by the formula

$$\tau_d + M \tau_s = \frac{\log_2 M}{R} \quad (1)$$

In systems which are average power limited (such as those using Nd:YAG lasers), the addition of dead time improves performance, by reducing the number of background counts per slot (since the slot time is reduced by  $\tau_d/M$ ). In systems which are peak power limited (e.g., those using semiconductor lasers), dead time degrades performance and is usually not used, i.e.,  $\tau_d = 0$  for these systems.

Item (18), the atmospheric transmission loss factor, includes transmission losses due to Earth or planetary atmospheres between the transmitter and receiver. It is assumed that both signal and background are subject to this loss. Item (19) is the average background radiance over the field of view of the detector. Table 2 gives a sample list of values for the operational parameters.

### III. Received Signal Power Calculation

Table 3 shows part of an optical link design control table (DCT) summarizing received signal power calculations for a system using the components and parameters described in Tables 1 and 2. In general, a DCT takes the form of an initial signal power (here the laser average power,  $P_\ell$ ), a number of multiplicative gain or loss factors leading to a received signal power, a required signal power (for the desired error rate), and a resulting system operating margin. The DCT is just a formal method for calculating the received signal via the range equation and comparing it to the signal required for some given error rate. Here, the range equation is given by

$$P_r = P_\ell g_t \eta_t \eta_{tp} \ell_s \ell_a g_r \eta_r \eta_\lambda \eta_D \quad (2)$$

where  $P_r$  is the received optical signal power,  $g_t$  is the transmitter telescope gain,  $\eta_{tp}$  is the transmitter pointing efficiency,

$\ell_s$  is the space loss factor,  $g_r$  is the receiver telescope gain,  $\eta_D$  is the detector truncation loss factor, and the remaining factors are defined among items (1)–(19) above. In sections IV to VI below we consider calculation of the various gain and loss factors which appear in Eq. (2). Section VII describes the received background power calculation.

## IV. Transmitter Factors

The first two factors in Eq. (2),  $g_t$  and  $\eta_t$ , together give the net on-axis gain of the transmitter optical system. The first factor,  $g_t$ , is the telescope gain, including effects of nonuniform aperture illumination, beam truncation, and a central beam obscuration, but not including transmission and reflection losses (see below). For telescopes with Gaussian-beam illumination and a central obscuration, the on-axis gain is given (Ref. 5) by

$$g_t = \left(\frac{\pi D_t}{\lambda}\right)^2 \frac{2}{\alpha_t} \left\{ e^{-\alpha_t^2} - e^{-\alpha_t^2 \gamma_t^2} \right\}^2 \quad (3)$$

where  $D_t$  is the transmitter aperture diameter,  $\gamma_t$  is the obscuration ratio (defined as the ratio of the central obscuration diameter to the main aperture diameter), and  $\alpha_t$  is the truncation ratio, given optimally (Ref. 5) by

$$\alpha_t = 1.12 - 1.30 \gamma_t^2 + 2.12 \gamma_t^4 \quad (4)$$

valid for  $\gamma_t \leq 0.4$ . The truncation ratio is the ratio of the main aperture diameter to the Gaussian-beam spot size. Equation (3) ignores the effect of the secondary element support struts, which for some telescopes can cause losses comparable to those due to the obscuration.

The transmitter optics efficiency,  $\eta_t$ , is given by item (5) above, i.e., it is to be specified as a basic component value. This term takes into account transmission and reflection losses in the transmitter, i.e., those in the relay optics, steering mirrors, and in the telescope. As noted above,  $\eta_t$  can in fact be readily calculated from the transmission and reflection coefficients of the individual optical components once the optical system has been specified in detail. For typical systems,  $\eta_t$  is in the range of 0.4 to 0.7.

The next factor in Eq. (2) is the transmitter pointing efficiency,  $\eta_{tp}$ . The pointing loss depends on the telescope gain (i.e., the narrowness of the transmitted optical beam) and on the statistics of the transmitter pointing system. Static errors in determining the desired pointing angle and in achieving that actual pointing angle lead to pointing bias errors; dynamic effects in measuring the desired angle and in maintaining a fixed pointing angle in the face of spacecraft base

motion (or apparent image motion in a turbulent atmosphere) lead to pointing jitter errors. Jitter errors are often assumed to be radially symmetric and Rayleigh distributed (Ref. 3). Quasi-static errors, such as “pointing control errors” which can occur in radio-frequency systems on deadband-controlled deep-space spacecraft, would normally be completely removed by the optical communications fine-pointing system. (For a known spacecraft, it would be possible to model some sources of base motion, such as attitude control jets, to produce a better estimate of optical system pointing errors.)

Pointing losses can be accounted for in several ways. Here, we assume that the mean value of the pointing loss factor  $\eta_{tp}$  is desired, and that dynamic errors are in fact Rayleigh distributed (Fig. 2 shows pointing error geometry). The overall pointing error (bias plus jitter) will then be given (Ref. 3) by the well-known Rice density so that

$$\eta_{tp} = \int_0^\infty \eta_{tp}(\phi) \frac{\phi}{\sigma_t^2} \exp \left\{ -\frac{\phi^2 + \epsilon_t^2}{2\sigma_t^2} \right\} I_o \left( \frac{\phi \epsilon_t}{\sigma_t^2} \right) d\phi \quad (5)$$

where  $\epsilon_t$  is the root-sum-square (RSS) two-axis pointing bias error,  $\sigma_t$  is the RSS two-axis jitter,  $I_o$  is the modified Bessel function of order zero, and  $\eta_{tp}(\phi)$  is the instantaneous pointing loss as a function of off-axis pointing angle,  $\phi$ . The term  $\eta_{tp}(\phi)$  is given for telescopes with Gaussian beam illumination and a central obscuration by the formula (Ref. 5)

$$\eta_{tp}(\phi) = \left[ \frac{\int_{\gamma_t}^1 e^{-\alpha_t^2 u^2} J_o \left( \pi \frac{D_t}{\lambda} \phi u \right) u du}{\int_{\gamma_t}^1 e^{-\alpha_t^2 u^2} u du} \right]^2 \quad (6)$$

where  $J_o$  is the Bessel function of order zero. (The dummy variable  $u$  in the integral is the radial position in the transmitter aperture.) For reasonable pointing errors, i.e., for  $\phi \leq \lambda/D_t$ ,  $\eta_{tp}(\phi)$  can be approximated (within 0.1 dB) by the series

$$\eta_{tp}(\phi) \cong \frac{1}{f_0^2(\gamma_t)} \left[ f_0(\gamma_t) + \frac{f_2(\gamma_t)}{2!} x^2 + \frac{f_4(\gamma_t)}{4!} x^4 + \frac{f_6(\gamma_t)}{6!} x^6 \right]^2 \quad (7)$$

where  $x = \pi(D_t/\lambda)\phi$  and the coefficients  $f_0$ ,  $f_2$ ,  $f_4$ , and  $f_6$  are given for several values of  $\gamma_t$  (and  $\alpha_t$  as given by Eq. 4) in Table 4.

Since performance can degrade rapidly with increasing pointing errors, it is sometimes preferable to use a burst-error method of determining pointing losses. In this case it is assumed that whenever pointing errors exceed some threshold value, a “burst error” occurs, i.e., that the link is essentially lost momentarily at the time. The pointing loss is calculated as the loss at the threshold value of the pointing error, so that acceptable communications performance is guaranteed except during the occasional “bursts.” In practice, an acceptable probability of burst error is chosen (e.g., one acceptable to the spatial tracking system and the synchronization systems), and a corresponding threshold pointing error (which is statistically exceeded at the burst error rate) is then calculated. The threshold pointing error then appears as a static (bias) error in the DCT, and jitter is ignored.

## V. Transmission Loss Factors

As in radio frequency systems, the largest loss in an optical communication system is usually the “space loss,” i.e., the loss associated with propagation through free space. The space loss factor,  $\ell_s$ , is given for all systems by

$$\ell_s = \left( \frac{\lambda}{4\pi L} \right)^2 \quad (8)$$

Due to the dependence on wavelength, the space loss incurred by optical systems is much larger (i.e., the factor  $\ell_s$  is much smaller) than in RF systems, but this is usually more than offset by higher optical antenna (telescope) gains.

Besides the space loss, there may be additional propagation losses if the signal passes through a lossy medium, e.g., a planetary atmosphere. Many optical links (like the one described in the Tables 1–3) are space-to-space links, i.e., they don’t enter an atmosphere and hence experience no such additional transmission losses. Models for determining optical losses due to transmission through the Earth’s atmosphere exist (see, for example, Ref. 6), but will not be considered here. Losses of 2 to 6 dB (depending on horizon angle) are typical for transmission from space to the Earth’s surface.

## VI. Receiver Factors

The next factor, receiver telescope gain,  $g_r$ , is the gain for an ideal receiving aperture with area equal to the unobscured part of the telescope, i.e., with area  $\pi D_r^2(1 - \gamma_r^2)/4$ . Hence  $g_r$  is given by

$$g_r = \left( \frac{\pi D_r}{\lambda} \right)^2 (1 - \gamma_r^2) \quad (9)$$

Transmission and reflection losses in the receiver are taken into account by  $\eta_r$ , the receiver optics efficiency. As in the transmitter, the optics throughput is assumed here to be a basic parameter, but in fact may be calculated from the knowledge of the optical layout in the receiver.

Although it can be considered as part of the overall receiver optical system, the narrow band filter transmission,  $\eta_\lambda$ , is usually treated as a separate factor. The narrow band filter is an important component in a direct-detection optical communications system, and its properties greatly affect the receiver’s sensitivity and rejection of background noise. Ideally, the filter should have 100% transmission in the passband and a very narrow (e.g., 1 angstrom or less) spectral bandwidth,  $\Delta\lambda$ . Practically, the achievable peak transmission is a function of the spectral bandwidth, with narrower filters usually having lower transmission factors. For this reason, performance is often maximized with filters having relatively wide bandwidths, e.g.,  $\geq 10$  angstroms.

The final factor in Eq. (2) is  $\eta_D$ , the detector truncation loss factor. Detector truncation loss occurs because not all of the light collected by the receiver aperture can be focused onto a finite-sized detector, especially when there may also be receiver pointing errors. The detector truncation loss is analogous to the transmitter pointing loss, and is given by

$$\eta_D = \int_0^\infty \eta_D(\phi) \frac{\phi}{\sigma_r^2} \exp \left\{ -\frac{\phi^2 + \epsilon_r^2}{2\sigma_r^2} \right\} I_o \left( \frac{\phi \epsilon_r}{\sigma_r^2} \right) d\phi \quad (10)$$

where  $\epsilon_r$  is the receiver RSS two-axis pointing bias error,  $\sigma_r$  is the RSS two-axis jitter, and  $\eta_D(\phi)$  is the instantaneous detector truncation loss as a function of off-axis pointing error, given by

$$\begin{aligned} \eta_D(\phi) &= \frac{2}{1 - \gamma_r^2} \\ &\times \frac{1}{2\pi} \int_0^{2\pi} \int_0^\theta \left[ J_1 \left( \pi \frac{D_r}{\lambda} \sqrt{\phi^2 + \xi^2 - 2\phi\xi \cos \psi} \right) \right. \\ &\quad \left. - \gamma_r J_1 \left( \pi \frac{\gamma_r D_r}{\lambda} \sqrt{\phi^2 + \xi^2 - 2\phi\xi \cos \psi} \right) \right]^2 \frac{d\xi}{\xi} d\psi \end{aligned} \quad (11)$$

where  $\theta$  is the receiver field-of-view,  $\gamma_r$  is the receiver obscuration ratio, and  $J_1$  is the Bessel function of order one. (Here the

integral is over the solid angle viewed by the detector.) Calculation of  $\eta_D$  using Eqs. (10) and (11) is rather difficult, requiring computation of a triple numerical integral. As a result, it is often assumed that  $\eta_D = 1$ . (Realistically,  $\eta_D \approx 1$  only in cases where no attempt is made to optimize background light rejection. In other cases  $\eta_D$  is in the range  $0.5 < \eta_D < 1$ .)

It is convenient to normalize the received power by the PPM word rate and by the energy per photon to obtain the number of received signal photons per PPM word (i.e., per laser pulse),  $N_s$ :

$$N_s = P_r \frac{\log_2 M}{R} \frac{\lambda}{hc}$$

Here  $hc = 1.986 \times 10^{-25}$  joule-meters, Planck's constant times the speed of light.

## VII. Received Background Power Calculation

In direct-detection systems, it is also necessary to determine the received background power (due to background radiance) in order to determine the required signal power and hence the operating margin. The received background power,  $P_b$ , is given by

$$P_b = B \lambda_a \frac{\pi D_r^2}{4} \left(1 - \gamma_r^2\right) \frac{\pi \theta^2}{4} \Delta \lambda \eta_r \eta_\lambda \quad (13)$$

where all the factors on the right-hand side have been defined above. Often the background consists of some point sources (e.g., stars) and some extended sources (e.g., planets, off-axis light scattered in the telescope, light scattered by an atmosphere), in which case the average radiance  $B$ , is given by

$$B = \frac{4}{\pi \theta^2} \sum_i A_i + \sum_i B_i \quad (14)$$

where  $A_i$  are the spectral irradiances (Watts/meter<sup>2</sup>/angstrom) of the point sources and  $B_i$  are the radiances (Watts/meter<sup>2</sup>/steradian/angstrom) of the extended sources. Table 5 gives a DCT-like summary of the background power calculation.

Noise power is most usefully normalized by the PPM slot rate, not the word rate, since in systems with dead time not all the collected background power in a word time is relevant. The number of received background photons per PPM slot time is given (using Eq. (1)) by

$$N_b = P_b \tau_s \frac{\lambda}{hc} \quad (15)$$

Systems using crystalline lasers (e.g., Nd:YAG) typically have fixed slot times in the range of 10 to 100 ns. Systems using semiconductor lasers typically have zero dead time, in which case the slot time is given (using Eq. (1)) by

$$\tau_s = \tau_s^{\max} = \frac{\log_2 M}{MR} \quad (16)$$

## VIII. Conclusion

In the preceding sections the optical signal and background power at the detector have been calculated from the transmitter laser output power, the background radiance, and a set of parameters that characterize the transmitter and receiver systems. The remaining steps in calculating communications performance involve only characterization of detector sensitivity, i.e., calculating the required signal power for a desired level of performance, and determining the resulting system margin.

In practice, communications link received power calculations must often be done repetitively. Besides the need to analyze many different potential optical link applications, it is usually necessary to design a given link iteratively in order to meet the desired requirements. For this reason, all the calculations described in this report (with the exception of Eqs. (10) and (11)), a set of predefined background sources (e.g., planets, stars, etc.), and required signal calculations for an "ideal" PPM receiver have all been incorporated into a computer program suitable for use on an IBM personal computer. The program generates an output equivalent to that in Tables 3 and 5, and can also automatically adjust any system parameter to fit a given margin. The program is an extremely useful tool for the design and analysis of optical communications links.

## References

1. Lesh, J. R., Marshall, W. K., and Katz, J., "A Simple Method for Designing or Analyzing an Optical Communication Link," *TDA Progress Report 42-85*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 25-31, 1986.
2. Dolinar, S. J., and Yuen, J. H., "A Note on Deep Space Optical Communication Link Parameters," *TDA Progress Report 42-70*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 105-109, 1982.
3. Vilnrotter, V. A., "The Effects of Pointing Errors on the Performance of Optical Communications Systems," *TDA Progress Report 42-63*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 136-146, 1981.
4. Yuen, J. H., ed., *Deep Space Telecommunications Systems Engineering*, Ch. 1, JPL Publication 82-76, Jet Propulsion Laboratory, Pasadena, Calif., also Plenum, 1983.
5. Klein, B. J., and Degnan, J. J., "Optical Antenna Gain, 1: Transmitting Antennas," *Appl. Opt.* 13, 9, Sept. 1974.
6. McCartney, E. J., *Optics of the Atmosphere*, John Wiley and Sons, New York, 1976.

**Table 1. Sample communications link component values**

Component	Value
1. Laser Power (Average)	0.200 Watts
2. Laser Wavelength	0.532 micrometers
3. Transmit Telescope Diameter	0.1 meters
4. Transmitter Obscuration Ratio	0.2
5. Transmitter Optics Efficiency	0.45
6a. Transmitter Pointing Bias Error	0.4 microradians
6b. Transmitter Pointing Jitter	0.8 microradians
7. Receive Telescope Diameter	1.0 meters
8. Receiver Obscuration Ratio	0.35
9. Receiver Optics Efficiency	0.7
10. Narrowband Filter Transmission	0.7
11a. Receiver Pointing Bias Error	0 microradians
11b. Receiver Pointing Jitter	0 microradians
12. Spectral Filter Bandwidth	10 Angstroms
13. Detector Diametrical Field of View	5 microradians

**Table 2. Sample communications link parameters**

Parameter	Value
14. Data Rate	30 kbps
15. PPM Alphabet Size	256
16. PPM Slot Time	10 nanoseconds
17. Link Range	$2.3 \times 10^{11}$ meters
18. Atmospheric Loss Factor	1.0
19. Background Radiance	0.2 W/m <sup>2</sup> /sr/A

**Table 3. Optical communications received signal power calculation summary**

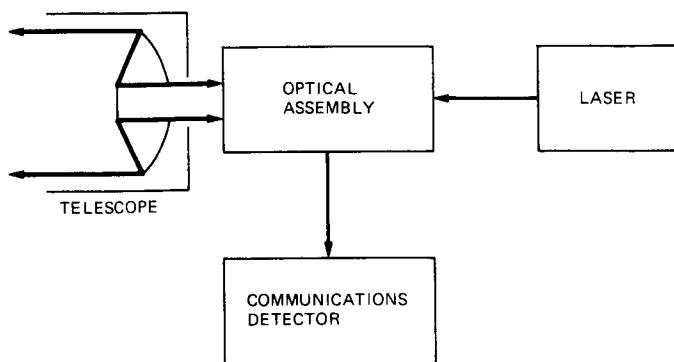
Received Signal Power	Factor	dB
Laser Power (Average)	0.200 W	23.0 dBm
Transmitter Telescope Gain 0.1 meter primary 0.02 meter obscuration $\lambda/D_t = 5.3 \mu\text{rad}$ .	$2.47 \times 10^{11}$	113.9
Transmitter Optics Efficiency	0.45	-3.5
Transmitter Pointing Efficiency 0.4 $\mu\text{rad}$ bias error 0.8 $\mu\text{rad}$ jitter	0.9	-0.5
Space Loss	$3.42 \times 10^{-38}$	-374.7
Atmospheric Loss	1.0	0.0
Receiver Telescope Gain 10 meter primary 3.5 meter obscuration	$3.06 \times 10^{13}$	134.9
Receiver Optics Efficiency	0.7	-1.5
Spectral Filter Transmission 10 A bandwidth	0.7	-1.5
Detector Truncation Loss	1.0	0.0
Received Signal Power	$1.03 \times 10^{-14}$ W	-109.9 dBm
Symbol Time	$2.67 \times 10^{-4}$ s	-35.7 dB/Hz
Photons/Joule	$2.68 \times 10^{19}/\text{J}$	154.3 dB/mJ
Received Signal Photons/Pulse	7.34	8.7 dB

**Table 4. Values of series coefficients for pointing loss calculation  
in Eq. (7)**

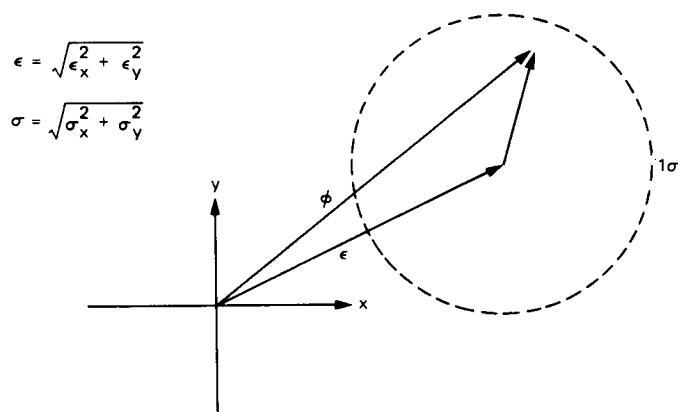
Transmitter Obscuration Ratio, $\gamma_t$	$f_0$	$f_2$	$f_4$	$f_6$
0.0	0.569797	-0.113421	0.0503535	-0.0292921
0.1	0.566373	-0.115327	0.0513655	-0.0299359
0.2	0.555645	-0.120457	0.0542465	-0.0317773
0.3	0.535571	-0.126992	0.0584271	-0.0344978
0.4	0.501381	-0.131777	0.0626752	-0.0374276

**Table 5. Received background power calculation summary**

Received Background Power	Factor
Background Radiance	0.02 W/m <sup>2</sup> /sr/Å
Atmospheric Loss Factor	1.0
Receiver Area	0.689 m <sup>2</sup>
1.0 meter primary	
0.35 meter obscuration	
Receiver Solid-Angle Field of View	$1.96 \times 10^{-11}$ sr
5.0 $\mu$ rad diameter FOV	
Spectral Filter Bandwidth	10 Å
Receiver Optics Efficiency	0.7
Spectral Filter Transmission	0.7
Received Background Power	$1.32 \times 10^{-12}$ W
Slot Time	$1 \times 10^{-8}$ s
Photons/Joule	$2.68 \times 10^{19}$ /J
Received Background Photons/Slot	0.354



**Fig. 1. Block diagram of an optical communications transceiver**



**Fig. 2. Pointing error geometry**

# A Cooled Avalanche Photodiode With High Photon Detection Probability

D. L. Robinson and B. D. Metscher  
 Communications Systems Research Section

*An avalanche photodiode has been operated as a photon-counting detector with 2 to 3 times the sensitivity of currently-available photomultiplier tubes. APD detection probabilities that exceed 27% and approach 50% have been measured at an optimum operating temperature which minimizes noise. The sources of noise and their dependence on operating temperature and bias voltage are discussed.*

## I. Introduction

The use of optical wavelengths for communication over planetary distances requires detectors sensitive enough to detect light pulses containing only a few photons. Presently the only detectors with high enough gain and low enough noise to be useful for detecting such faint signals are photomultiplier tubes (PMTs); however, their low quantum efficiency, the fragility of vacuum tube packages, and the high anode voltages required are important disadvantages for a space-based optical receiver. As a possible alternative, avalanche photodiodes (APDs) have been tested in a single photon detection mode. When cooled to reduce the number of thermal carriers, an APD can be reverse-biased beyond its breakdown voltage to give it a very high ( $10^7$  to  $10^8$ ) internal gain. APDs have a higher quantum efficiency and the reliability of solid state devices; thus they may be a useful alternative to PMTs.

The theory of operation, statistical models of signal detection and noise generation, and quenching systems have been investigated by others (Refs. 1-6). The present work centers

on understanding the operational characteristics of APDs in a photon counting mode, and in particular on measuring the probability of photon detection as a function of temperature and bias voltage.

The theory of operation of APDs, the sources of noise, and the problem of quenching have been discussed in previous work (Ref. 7). The experimental procedures used to examine the characteristics of APDs and to measure detection probabilities are described here. Experimental results are then presented and discussed.

## II. Experimental Setup and Procedure

A schematic of the experimental setup is shown in Fig. 1. The APD is cooled in an MMR Technologies K7701 System I microminiature refrigerator, in which the diode is mounted on a cold finger inside a small evacuated chamber and connected to the quenching circuit by manganin wire leads. The light source is a Mitsubishi ML-3101 laser diode emitting at 822 nm at 20 mA operating current. The laser may be operated as

either a pulsed or a continuous source. A calibrated beam-splitter directs a known fraction of the laser light to a United Detector Technology S550 optical power meter to monitor the light level incident on the detector being tested. The light intensity is attenuated to a few photons per laser pulse by neutral density filters calibrated at 822 nm, then a second beamsplitter divides the beam between the APD and PMT. Two different PMTs were used, an RCA model C31034 and an RCA C31034A-02, both with GaAs photocathodes. These were cooled to 240 K in a TE refrigerator housing and were operated at 1600 V. All optical components were contained in a light-tight enclosure. Data collection and reduction were accomplished using a pair of HP 5370A time interval counters, a Tektronix 7834 storage oscilloscope, and an IBM PC AT (Fig. 2).

The APDs tested were RCA type C30902S IR-sensitive silicon diodes. This model is similar to the one tested in previous work (Ref. 7; RCA C30817), but the C30902 diodes are selected by the manufacturer for performance as photon counters and are sold as "optimized" for operation in a photon counting mode.

The APDs were tested using a passive quenching circuit shown schematically in Fig. 3. The recovery time of the detector using this circuit, i.e., the time between a detection event and restoration of the bias voltage across the diode to its original value, was approximately 6  $\mu$ s.

An active quenching circuit devised by Cova, et al., (Ref. 5) was built and tested, but effective implementation of their discrete-component, high-frequency feedback circuit proved impractical. However, since active quenching can reduce the detector's dead time to as little as 10 ns, further consideration of active quenching circuits is planned.

The frequency of dark noise counts was measured for each diode as a function of temperature and reverse bias voltage, and dark noise was measured for the PMT. These data were collected by simply counting noise events with a time interval counter and reading and storing the count data with an IBM PC.

To measure the APD detection probabilities, the light source was pulsed at 15 kHz with pulses of about 50 ns duration. The voltage pulse from the laser power source was used to gate a logic circuit (Fig. 4) which transmits either the APD output counts occurring during a laser pulse ("true" counts), those occurring outside a laser pulse ("false" counts), or all output counts.

The PMT measured the beam intensity after attenuation. From this measurement, using the quantum efficiency of the PMT and the measured efficiency of the beamsplitter and

other optics, the light intensity incident on the APD was calculated. Detection probabilities were calculated from this incident intensity and the measured number of APD detection events.

### III. Results and Discussion

#### A. Dark Noise

Figure 5 summarizes the data for dark count frequencies for one RCA APD. Measurements were made at temperatures between 77 K and room temperature. As can be seen in these plots, the dark noise from the APDs increased more slowly with increasing bias voltage at intermediate temperatures than it did at higher and lower temperatures. The optimum operating temperature for these devices appears to be about 200 K. At this temperature, the dark noise counts were less than 100 per second for a bias voltage of more than 10 V above the breakdown voltage. Operating the detector at the highest possible bias voltage is desirable in order to obtain the greatest sensitivity.

The statistical frequency distribution of the lengths of the time intervals between noise counts was compared with a Poisson distribution using the chi-squared goodness-of-fit test. It was found that the dark noise counts were well-approximated by a Poisson process at very low (near 80 K) and high (near 300 K) temperatures. At intermediate temperatures (between 120 K and 200 K) the chi-squared test failed to show a good fit of the dark noise distribution to a Poisson distribution.

We interpret these results as being dependent on the dead time of the passive quenching circuit and may differ with the use of an active quenching circuit where dead times are approximately 10–20 ns. At high temperatures ( $> 200$  K) the thermal carriers effectively dominate (due to the circuit dead time) while at lower temperatures the dominant noise is due to trapped carriers or secondary avalanches. Trapping of carriers occurs during an avalanche, presumably at imperfection sites in the semiconductor crystal lattice. Carriers trapped after a detection or noise event can remain in the lattice and dislodge, initiating a secondary avalanche, after a time ranging from nanoseconds to tens of minutes. Both the thermal generation of noise carriers and the release of trapped carriers are expected to be Poisson processes.

At high temperatures (i.e., near room temperature), since electrons are more mobile in the crystal lattice, trapped carriers are dislodged more quickly and most initiate secondary avalanches within the 6  $\mu$ s dead time of the APD quenching circuit. Thus only a few avalanches due to trapped carriers are detected. At low temperatures (80 K to 120 K), trapped car-

riers remain trapped longer, and noise events due to them dominate those due to thermal carriers, whose numbers are greatly reduced at low temperatures.

At intermediate temperatures, detected events include a significant proportion of both the Poisson-distributed primary (thermal) events and the conditionally Poisson-distributed secondary (trapping) events. Thus the noise at intermediate temperatures does not fit a simple Poisson distribution.

## B. Detection Probabilities

The probability of single photon detection was taken as the ratio of the number of photons detected per second by the APD to the number of photons per second incident on its active region. The intensity of incident light was determined from two independent measurements, one by the UDT photodiode power meter, the other by the PMT. From the intensity of the beam at each of these two points, using the measured transmission and reflection efficiencies of the two beamsplitters and the measured transmittance of the filters and other optics, the number of photons falling on the APD was calculated. Incident light intensities on the APD calculated from the photodiode data agreed within approximately 30% (or about 1/3 photon since we are detecting single photons) of that calculated from the PMT output data. As a further verification of the incident light intensity, a second PMT was used. The two PMT count rates were within approximately 25% agreement for similar incident light intensities of a few photons. This discrepancy reflects the uncertainties in the manufacturer's specifications for the quantum efficiencies of the PMTs. Despite this, we are reasonably confident that the light levels incident on the detectors are equal to or less than one photon per pulse for the experiment described here.

The output count frequency measurements of the "true," "false" and dark APD counts and the PMT output counts are summarized graphically in Fig. 6 for one RCA APD. Results have been scaled to represent a constant light intensity to both the PMT and APD. (Experimentally, approximately 1.7 times more light was incident on the PMT than on the APD.)

As can be seen in Fig. 6, dark counts remain less than 100 counts per second even with an over-bias of 10 volts (as described earlier). When light pulses are incident on the APD,

more trapped carriers are generated. These carriers detrap causing a secondary avalanche after the primary avalanche has occurred. These secondary avalanches or false pulses increase the effective dark count rate. The rate of occurrence of these false counts is dependent on the over-bias of the APD. As the over-bias is increased, the false count rate also increases. It is expected that the false count rate is also a function of the light intensity incident on the APD. This will be further investigated at a later time. As can be seen in Fig. 6, the APD is still quite effective in spite of the generation of secondary avalanches since false count rates remain below 100 counts per second for an over-bias of about 1.5 volts on the APD. Even at this low over-bias, the APD true detection count rate exceeds the PMT count rate as shown in Fig. 6. As expected, the true count detection rate increases with over-bias voltage.

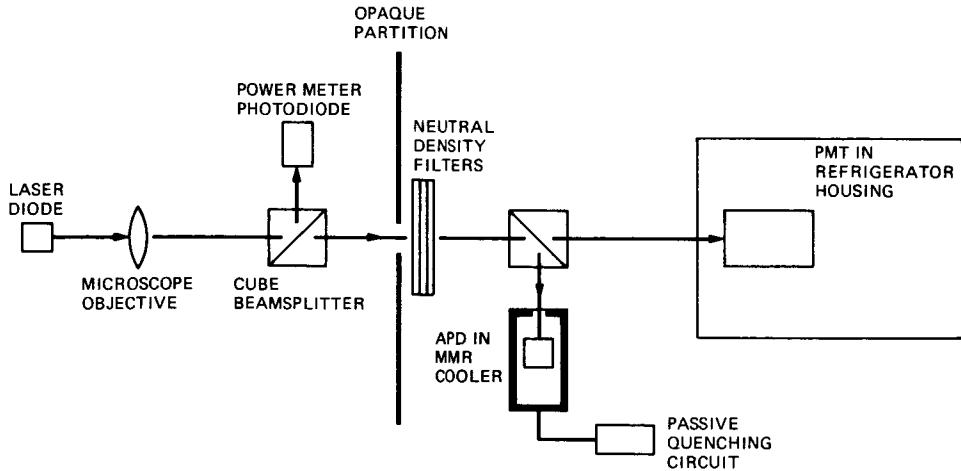
From this true count detection rate, a single-photon detection probability was calculated (as described above) for each voltage tested. Results are summarized in Fig. 7, and the appropriate false count levels are indicated. The uncertainty in this curve is approximately 30%, due mainly to uncertainties in the sensitivities of the PMTs. For a false count rate or effective dark count rate of 100 counts per second and an over-bias of about 1.5 volts, single-photon detection probabilities of approximately 27% were achieved. This is over twice that achievable with the C31034 PMT (quantum efficiency  $\sim 12\%$  at 822 nm, cooled to 240 K). As can be seen from Fig. 7, single-photon detection probability approaches 50% for a 10 volt over-bias; however, the effective dark count rate nearly equals the signal count rate.

## IV. Conclusion

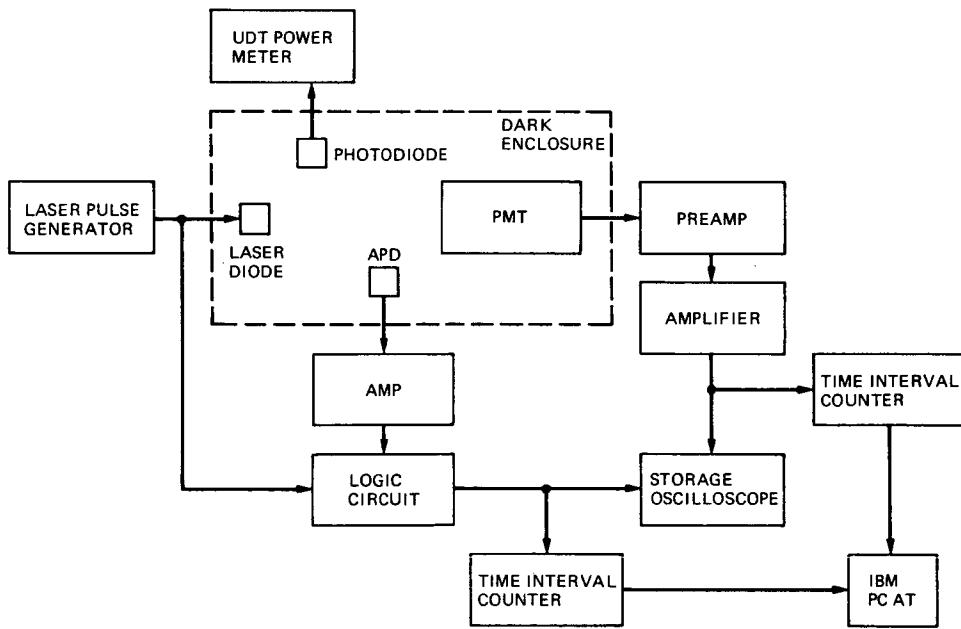
Commercial avalanche photodiodes have been operated as single photon detectors by optimizing their operating temperatures and bias voltages. When operated under these conditions, the sensitivity of an APD can exceed that of a photomultiplier tube by more than a factor of two with less than 100 noise counts per second. At this sensitivity, the APD probability of detection was approximately  $27 \pm 8\%$  at an over-bias of 1.5 volts. Thus APDs represent an attractive alternative to PMTs for applications in deep-space optical communications, as well as in other applications requiring sensitive, reliable, small or lightweight detectors, such as astronomy, remote sensing and fiberoptic communications.

## References

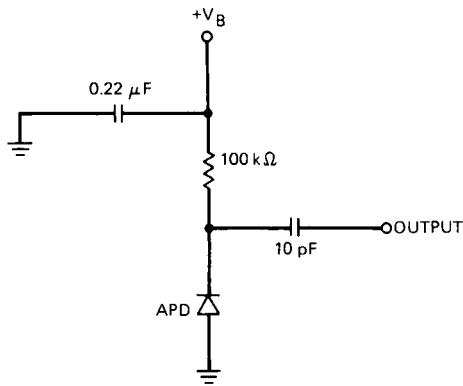
1. McIntyre, R. J., The Distribution of Gains in Uniformly Multiplying Avalanche Photodiodes: Theory, *IEEE Trans. Electron. Dev.*, ED-19, pp. 703-713, 1972.
2. McIntyre, R. J., On the Avalanche Initiation Probability of Avalanche Diodes above the Breakdown Voltage, *IEEE Trans. Electron. Dev.*, ED-20, No. 7, pp. 637-641, July 1973.
3. Estrom, P. A., Triggered-Avalanche Detection of Optical Photons, *J. Appl. Phys.* 52, pp. 6974-6979, 1981.
4. Oldham, W. G., Samuelson, R. R., and Antognetti, P., Triggering Phenomena in Avalanche Diodes, *IEEE Trans. Electron. Dev.*, ED-19, pp. 1056-1060, 1972.
5. Cova, S., Longoni, A., and Ripamonti, G., Active-quenching and Gating Circuits for Single-photon Avalanche Diodes (SPADs), *IEEE Trans. Nucl. Sci.*, NS-29, pp. 599-601, 1982.
6. Tan, H. H. Avalanche Photodiode Statistics in Triggered Avalanche Detection Mode, *TDA Progress Report 43-79*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 69-80, July-September 1984.
7. Robinson, D. L., and Hays, D. A., Photon Detection with Cooled Avalanche Photodiodes: Theory and Preliminary Experimental Results, *TDA Progress Report 42-81*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 9-16, January-March 1985.



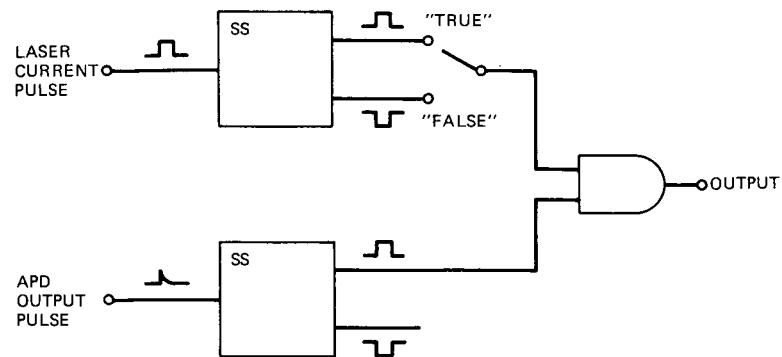
**Fig. 1. Arrangement of optical components inside dark enclosure**



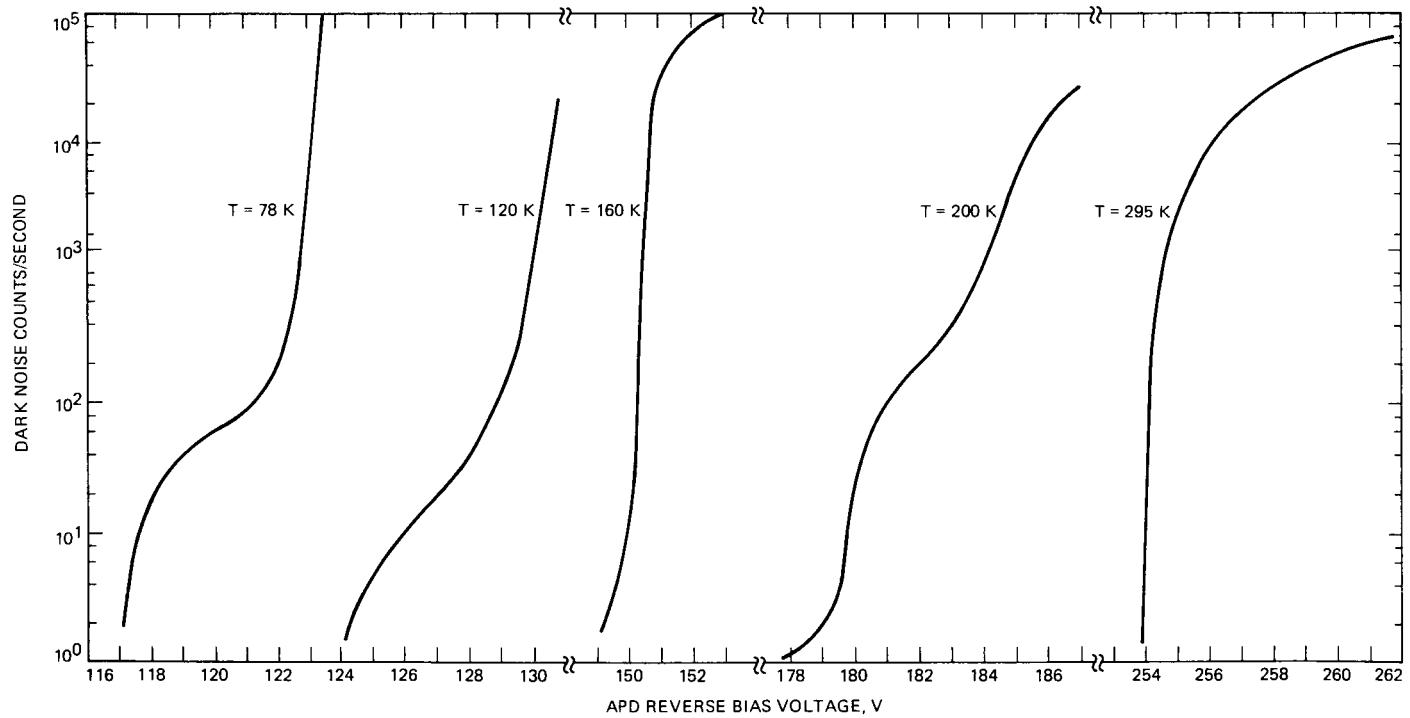
**Fig. 2. Schematic representation of APD detection probability data acquisition and reduction apparatus. The amplifier for the APD output signal is necessary only for input to the TTL logic circuit. In general, further amplification of an APD signal pulse is not needed.**



**Fig. 3. Passive quenching circuit. Dead time of the detector using this circuit is about 6  $\mu$ s.**



**Fig. 4. Logic circuit for determining coincidence of APD output pulses with laser input pulses**



**Fig. 5. Measured frequency of dark noise counts as a function of APD bias voltage and temperature (horizontal axis is discontinuous)**

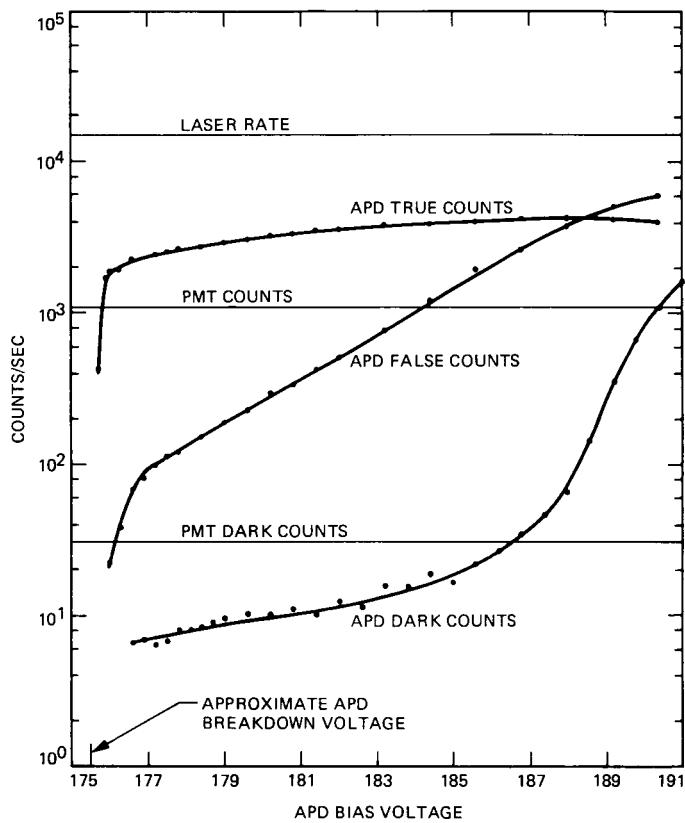


Fig. 6. APD output count frequencies as functions of bias voltage at 200 K for an RCA 30902S APD. Laser pulse rate was 15 kHz. The PMT (RCA 31034, quantum efficiency  $\sim 12\%$ ) count rate has been scaled to the same incident light intensity as the APD. APD breakdown voltage was approximately 175.5 V at 200 K.

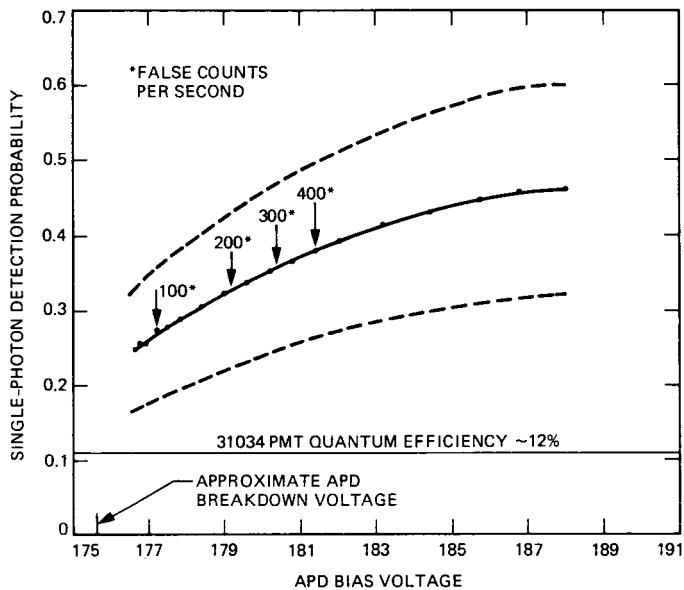


Fig. 7. APD single-photon detection probability as a function of bias voltage at 200 K. Effective dark ("false") count rates are indicated. Uncertainty (dashed curves) in this curve is estimated to be 30%.

26-32

Q.S.

N87-15335

TDA Progress Report 42-87

July-September 1986

# Fraunhofer Filters to Reduce Solar Background for Optical Communications

E. L. Kerr

Communications Systems Research Section

*A wavelength that lies within a spectral interval of reduced solar emission (a Fraunhofer line) can carry optical communications with reduced interference from direct or reflected background sunlight. Suitable Fraunhofer lines are located within the tuning range of good candidate lasers. The laser should be tunable dynamically to track Doppler shifts in the sunlight incident on any solar system body that may appear in the background as viewed by the receiver. A Fraunhofer filter used with a direct-detection receiver should be tuned to match the Doppler shifts of the source and background. The required tuning calculated here for various situations is also required if, instead, one uses a heterodyne receiver with limited post-detection bandwidth.*

## I. Introduction

### A. Sunlight Interference

Sunlight interferes with optical communications when it enters the receiver telescope and detectors. It may be scattered into the receiver from the blue sky, or it may enter as part of the background when the telescope is looking at the laser source. The sun itself may be in the background. If, instead, a planet or a moon appears in the background, the interfering light is still sunlight after reflection and scattering from the background surface.

### B. Lines in the Solar Spectrum

No matter how the sunlight enters the receiver, it still has the spectrum of sunlight. In particular, there are many narrow wavelength intervals in which the sunlight is greatly reduced because of a natural filter existing over the sun. This filter

consists of iron and other vaporized materials in the solar atmosphere above the photosphere (the visible surface of the sun). The region where these materials exist is called the "reversing layer" because the materials absorb rather than emit light at characteristic wavelengths, producing a spectrum that is the reverse of an emission spectrum. Each of the resulting intervals of low sunlight emission is called a "Fraunhofer line."

### C. Discriminating Against Sunlight

A Fraunhofer filter is a very narrow optical filter that only admits light whose wavelength lies within a Fraunhofer line. Viewed through a Fraunhofer filter during the day, the sun, the sky and the ground look dark; at night all the planets and their moons look dark. A laser whose wavelength falls within the Fraunhofer line looks like one bright spot in the darkness, by day or by night.

A closer look through the filter shows that the limbs of the sun and the Jovian planets are bright near the equators, because of appreciable Doppler shifts caused by the high surface rotation velocities of those bodies. Also, at certain times the planet Mercury looks bright as it moves rapidly toward or away from the sun in its highly elliptical orbit. To remain visible, a laser that moves through the solar system should tune itself to the center of the Fraunhofer line as reflected by any object in the background. The Fraunhofer filter would have to be tuned accordingly.

As the earth moves in its orbit it also approaches or recedes from all these bodies. Therefore, as one looks through the filter and turns from viewing the sun to viewing other solar system bodies, one would need to tune the Fraunhofer filter to compensate for additional Doppler shifts.

#### D. Application

On any mission employing an optical data link to the earth, the chief interference may be sunlight, direct or reflected from a planet or satellite. An ideal sunfilter with a highly transmitting passband just wide enough to admit the laser signal would reduce solar background to tolerable levels. Sunlight admitted within the passband would continue to interfere, however.

If the laser wavelength were chosen to operate close to or within a Fraunhofer line, sunlight interference would be reduced further and communication would be improved. The greatest improvement would be realized by tuning the source laser to the minimum of a deep Fraunhofer line and narrowing the filter passband to lie entirely within the Fraunhofer line.

In reality the passband must either be wide enough to allow for all Doppler shifts to be encountered during the mission, or it must be tunable (or at least selectable) as the received signal wavelength changes.

There is also a trade-off in filter design between narrowness of the passband (to reduce background interference) and peak transmission (to maximize the received signal). It is not enough simply to maximize the signal-to-noise ratio; see Ref. 1.

#### E. Simplest Application

An optical communications system designer may choose a nontunable source laser for stability, reliability, and other reasons. The wavelength chosen for such reasons is unlikely to coincide exactly with a Fraunhofer line. Also, the receiver filter passband may be chosen as wide as 1 or 2 nm (10 or 20 Å) in order to obtain transmission of 70% to 80%. If these choices are made for a mission to an outer planet, the back-

ground of reflected sunlight when the planet fills the receiver field-of-view may be tolerable.

Even with these choices, however, the center frequency of the filter passband should still be chosen to take advantage of the existence of Fraunhofer lines. After all expected Doppler shifts of the received signal have been included, the filter passband should be chosen to include as many strong Fraunhofer lines as possible. Each included line could reduce background sunlight interference by 5% to 10%.

## II. Fraunhofer Lines and Filters

### A. Fraunhofer Lines

A high-resolution solar spectrum shows hundreds of dark lines. In some of them solar emission is reduced to 10% or less. Typical widths are 0.01 nm (0.1 Å). The center of the lines might be found to within an uncertainty of about 0.001 nm or 0.01 Å. (Since that uncertainty corresponds to 1 pm, I have chosen to state linewidths and Doppler shifts in picometers for convenience in this article. Sub-picometer shifts are insignificant.) Some lines lie close together, making broader dark bands. The lines are mostly in the visible region. The ones that correspond to absorption in the solar atmosphere are called Fraunhofer lines. The absorption is mainly by neutral or ionized elements, especially iron vapor.

The solar spectrum also contains many lines due to earth atmosphere absorption. These lines are of no help in reducing background, since they would also reduce signals transmitted through the atmosphere. They are distinguishable from solar lines because of the temperatures at which the absorbing molecules could exist. Water vapor, carbon dioxide, and diatomic oxygen are found in the earth atmosphere; solar absorbers such as iron vapor are mostly atomic. Solar molecules contain at most a few strongly bound atoms.

A detailed solar spectrum appears in Ref. 2. A sample is shown in Fig. 1, covering 529.1 nm to 532.4 nm. Most of the lines in the solar spectrum have been attributed to various molecules and radicals in Ref. 3. Consultation with that reference is necessary to be sure that the lines are Fraunhofer lines and not absorption lines in the earth's atmosphere.

One candidate wavelength for optical space communications is half that of a Nd:YAG laser, because of that laser's highly efficient use of power even after frequency doubling (Ref. 4). The halved wavelength with the next-to-lowest threshold corresponds very closely to the Fraunhofer line listed in Ref. 3 as centered at 530.7369 nm with an 8.6 pm width. The strongest cw laser wavelength, 1064.1 nm at room temperature, when halved falls into a region without Fraun-

hofer lines. However, if the YAG temperature is raised or lowered by about 20°C, the halved wavelength will fall either within the pair of singly ionized iron lines near 531.7 nm or the neutral iron line at 532.4 nm.

## B. Fraunhofer Filter

A Fraunhofer filter consists of a multi-layer dielectric interference filter and a Fabry-Perot etalon. The theory of Fraunhofer filter design is reviewed in Ref. 5 and a practical example is described in Ref. 6.

By itself, the multi-layer dielectric filter may have a passband as narrow as 1 to 2 nm. Transmission may still be as good as 70%. Further attempts to narrow the passband of a dielectric filter would pay a high cost in reduced transmission.

The Fabry-Perot etalon consists of two parallel, flat, highly reflecting optical surfaces separated a distance  $d$  by a medium of refractive index  $n$ . Perpendicularly transmitted wavelengths are those which are equal to  $nd$  divided by any natural number. Transmission drops very rapidly as the wavelength deviates from a transmitted wavelength. As the wavelength continues to vary, it approaches the next transmitted wavelength and the transmission rises rapidly again.

Even a thin etalon has many passbands. The finesse (defined as the width-to-spacing ratio of the passbands) depends on the geometric mean  $R$  of the reflectivities  $R_1$  and  $R_2$  of the two surfaces ( $R^2 = R_1 R_2$ ). This finesse may be degraded by surface imperfections, surface misalignment, and absorption on or between the surfaces.

When the multi-layer dielectric filter and the etalon are combined to make a Fraunhofer filter, the etalon spacing is chosen so that one of its passbands corresponds in center frequency and width to the desired Fraunhofer line. The multi-layer dielectric filter passband is also centered on the Fraunhofer line; its width is chosen only narrow enough to block the neighboring etalon passbands. The peak transmission of the multi-layer dielectric filter can be relatively high since its passband need not be very narrow. Overall transmission of a Fraunhofer filter may be 30% to 40%.

## C. Tuning Methods

A Fraunhofer filter is tunable over a narrow range by varying the etalon refractive index or the spacing, or both. A solid glass etalon could be temperature-tuned by choosing a material such as Pyrex with a large coefficient of thermal expansion. Then  $d$  would increase with increasing temperature. Many glasses have large positive values of the rate of change  $dn/dT$  of refractive index with absolute temperature  $T$ . A glass combining a large coefficient of thermal expansion

with a large positive  $dn/dT$  gives the greatest tuning sensitivity to temperature.

An etalon made from two optical flats with an air space between could be tuned sufficiently by fractional atmosphere pressure changes. The spacer thickness could be varied thermally or piezoelectrically.

Depending on the light entrance conditions, a designer might wish to hold the temperature, pressure, and spacing constant and tune the filter by tilting it.

## III. Adaptive Compensation for Three Doppler Shifts

There are four objects involved in the analysis of background interference in an optical data link: the sun ( $\odot$ ), the source ( $s$ ), the body that appears in the background ( $b$ ) of the source, and the receiver ( $r$ ). Radius vectors from one of these objects to another may be distinguished by an ordered pair of subscripts indicating the tail and head of the vector. The relative line-of-sight velocity of body 2 with respect to body 1 is the component of the relative velocity along the line of sight:

$$v_{12} = \bar{r}_{12} \cdot \dot{\bar{r}}_{12} / (\bar{r}_{12} \cdot \bar{r}_{12})^{1/2}$$

The three velocity components that cause Doppler shifts are the sun-background velocity  $v_{\odot b}$ , the background-receiver velocity  $v_{br}$ , and the source-receiver velocity  $v_{sr}$ . In the vicinity of the solar system and on typical missions all these velocities are small compared with the speed of light  $c$ .

A signal frequency shift  $\Delta f_{sr}$  or a signal wavelength shift  $\Delta\lambda_{sr}$  due to the Doppler effect and the line-of-sight velocity component  $v_{sr}$  between the source and the receiver is

$$\Delta f_{sr} = f_{sr} v/c$$

Also

$$(\Delta f_{sr})/f = -(\Delta\lambda_{sr})/\lambda$$

Other shifts are calculated with similar formulas using the appropriate subscripts.

When the background interference is reflected or scattered sunlight, the Fraunhofer lines of the sun will be reproduced in the spectrum of the background (provided the illuminated surface has no significant fluorescence). The wavelengths of

the Fraunhofer lines in the incident light will be shifted up or down as the surface is moving toward or away from the sun. These shifts will be referred to as "incident Doppler shifts."

When the source is seen against a background, the wavelengths of the Fraunhofer lines in the background will suffer a second Doppler shift if the background surface is moving toward or away from the receiver. These shifts will be referred to as "reflected Doppler shifts."

The laser source wavelength will be Doppler-shifted up or down as the source approaches or recedes from the receiver. These shifts will be referred to as "source Doppler shifts."

The two background shifts, incident and reflected, add algebraically. The dependence of the sum on the velocity vectors between the sun and the surface, and between the surface and the receiver, is complicated. The source shift may not correspond to either of the background shifts or to their algebraic sum. Hence, if the source laser wavelength is to stay within a Fraunhofer line of the background, it will have to be tunable as the background changes. Tuning has the effect of making small color changes; thus the source laser must act like a chameleon, except that the reason for the color change is opposite to that of the chameleon. The chameleon changes color to make itself invisible against the background; the source laser changes color to make its signal visible.

The source and reflected background shifts will be identical unless the source and the background surface have different velocity components toward the receiver.

For instance, if a space probe has made a soft landing on a planetary or satellite surface, it shares the motion of the surface. Sunlight falling on the surface may be Doppler-shifted relative to the original solar wavelengths. If the source laser makes the appropriate chameleon tuning shift, the source wavelength will remain within the line whatever the motion of the receiver relative to the probe and surface may be. The receiver would merely tune its filter to the source wavelength it saw.

If the source is flying over the background, then its velocity component  $v_{sr}$  is in general different from the velocity component  $v_{br}$  of the background surface, both as measured from the receiver. The required chameleon tuning shift is  $(1 + v_{br}/c)/(1 + v_{sr}/c) - 1$ . Since all the speeds are small compared with the speed of light, this tuning shift is approximately  $(v_{br} - v_{sr})/c = v_{bs}/c$ . This means it would be sufficient for the source laser to tune itself to what it perceives as the center frequency of the background Fraunhofer line when measured from the spacecraft. It need not consider the velocity of the receiver.

It only needs to know the direction of the receiver in order to look in the opposite direction and find out what the background is.

The receiver behavior is simpler. It needs only to tune to the received source wavelength. It may assume the source is centering its wavelength on the Fraunhofer line.

## IV. Doppler Shifts for Typical Situations

For yellow light (500 nm, 5000 Å, 600 THz) a shift of 100 pm (0.1 nm, 1 Å, 120 GHz) corresponds to  $v = 60$  km/s. A shift of 10 pm or 12 GHz, comparable with a single Fraunhofer linewidth, corresponds to  $v = 6$  km/s. Shifts smaller than 1 pm would be negligible.

Solar system velocities arranged from largest to smallest are generally due to planetary revolution about the sun, satellite orbital revolution about a planet, and surface rotation of a planet. Many of these velocities are tabulated in Ref. 7. The effects of these will be considered first for the incident light, and then for the background and source shifts relative to the receiver.

### A. Incident Shifts

Doppler shifts in the illumination from the sun would be zero if the planets were in perfectly circular orbits. In fact, all the orbits exhibit some ellipticity. At the apses (perihelion and aphelion) the Doppler shift is zero. The maximum Doppler shift occurs when

$$\cos E = e$$

where  $E$  is the eccentric anomaly, and  $e$  is the eccentricity. The maximum speed away from the sun is

$$v_{\odot b} = 2\pi ae/[P(1 - e^2)^{1/2}]$$

where  $P$  is the period of revolution, and  $a$  is the orbital semi-major axis. These speeds and the corresponding Doppler shifts are displayed in Table 1.

### B. Background Shifts

Angular rotation rates follow no particular pattern in the solar system. Surface rotation velocities are the product of the angular rotation rate and the radius of the body, so they are large only for the largest bodies. The sun itself shows rotational Doppler shifts of its Fraunhofer lines near the equatorial limbs. No natural satellite and no terrestrial-like planet has a surface rotation velocity large enough to cause an

appreciable Doppler shift. Equatorial velocities and corresponding Doppler shifts are shown in Table 2.

Sunlight falling on these planets would be shifted toward shorter wavelengths from the approaching limb, and toward longer wavelengths from the receding limb. Therefore, these incident Doppler shifts would only require an additional chameleon shift as the probe appeared to cross one of the planet limbs close to the equator.

If the probe made a soft landing on one of these planets near the equator, the required chameleon shift would be largest at local dawn or dusk. However, the apparent surface for a Jovian planet is really the top of the clouds. Even a balloon probe would rapidly sink out of sight of earth beneath them. Therefore, one need not usually consider a soft landing on any body large enough and rotating rapidly enough to have a significant rotational Doppler shift.

Objects too small to retain much atmosphere (Mars, Mercury, Pluto, natural satellites, or the asteroids) are usually too small to have appreciable surface velocity Doppler shifts, even if their rotational rates are large.

At times when a planet nears occultation by the sun, the radial component of its velocity relative to earth drops nearly to zero. Therefore, when a space probe is near or on a planet, and the line-of-sight to earth passes close to the sun, both the sun glare and the planetary background can be blocked by a Fraunhofer filter.

### C. Source-Receiver Shifts

A space probe may have any velocity relative to the earth. However, it will seldom have another body as a background unless it is close to that body. It will stay close only if it has approximately the same velocity as the body. It is, therefore, nearly sufficient to analyze the velocities of the planets relative to the earth.

Planetary revolution velocities decrease inversely as the square root of the radius from the sun. The formula follows from Kepler's law:

$$\begin{aligned} v_{\text{planet}} &= 2\pi (1 \text{ AU})^{3/2} \text{ yr}^{-1} / (r_{\text{sun-planet}})^{1/2} \\ &= (30 \text{ km/s}) (1 \text{ AU}/r_{\text{sun-planet}})^{1/2} \end{aligned}$$

The same formula applies for satellite orbital revolution about a planet, whether it is the receiver in orbit about the earth, or the signal source on a spacecraft in orbit about the target planet. The constant depends on the planet mass.

Using the universal gravitational constant  $G = 6.670 \times 10^{-11} \text{ N}\cdot\text{m}^2\cdot\text{kg}^{-2}$ ,

$$v_{\text{satellite}} = (G M_{\text{planet}} / r_{\text{planet-satellite}})^{1/2}$$

The earth's revolution velocity causes almost the largest and also the most troublesome Doppler shift of all those that must be considered for optical communications. Fortunately, it is also nearly the slowest. The repetition cycle time is one year. Actually the repetition cycle time depends on the periods of both planets at the ends of the data link, but mostly it depends on the period of the inner planet. There are only two planets with shorter periods than earth's. The shortest repetition cycle time for a revolution Doppler shift is that of Mercury, 88 days. Rapid tuning of the filter is, therefore, not required for most projected missions.

The maximum radial component of the relative velocity between an inner and outer planet in circular orbits in the solar system is given by

$$v = v_{\text{in}} [1 - (r_{\text{in}}/r_{\text{out}})^{3/2}]$$

and the minimum is  $-v$ . The maximum occurs when the inner planet lags behind the outer planet by an angle whose cosine is  $r_{\text{in}}/r_{\text{out}}$ , and the minimum occurs when the inner planet leads by the same angle. See Table 3 for these velocities and Doppler shifts.

If the near-earth end is on a satellite, there are Doppler shifts due to the earth-orbital velocities. These shifts decrease with increasing satellite altitude. If tuning is required, its repetition cycle time corresponds to the orbital period. At the probe end, if the source is in an isosynchronous orbit, there can be appreciable Doppler shifts for the rapidly spinning planets beyond the earth. Refer to Table 4 for these velocities and Doppler shifts.

The surface escape velocity is just the square root of 2 times the orbital velocity of a satellite at negligible altitude. (Note that the Doppler shift for a satellite in low earth orbit is just the table entry for earth divided by the square root of 2.) These velocities are given for major bodies in Table 5. The Doppler shifts caused by these velocities affect optical communications only during the descent or return phases of a mission to the surface of a planet.

### V. Conclusions

Fraunhofer filters can improve optical space communications. It is possible to find a close match between a Fraunhofer line and a useful laser wavelength. Laser wavelengths

and Fraunhofer filters can be tuned to match shifts in the Fraunhofer lines due to the Doppler effect. Typical Doppler shifts are less than 50 pm (0.5 Å) or 60 GHz for almost all missions outside the orbit of Mercury at almost all times.

Doppler shifts have been calculated for many situations to aid designers who would use either a Fraunhofer filter with direct detection or heterodyne detection with a reasonably narrow post-detection bandwidth.

## References

1. Lesh, J. R., Marshall, W. K., and Katz, J., "A Simple Method for Designing or Analyzing an Optical Communication Link," *TDA Progress Report 42-85*, pp. 25-31, Jet Propulsion Laboratory, Pasadena, CA, May 15, 1986.
2. Minnaert, M., Mulders, G. F. W., and Houtgast, J., *Photometric Atlas of the Solar Spectrum*, D. Schnabel, Amsterdam, 1940.
3. Moore, C. E., Minnaert, M. G. J., and Houtgast, J., *The Solar Spectrum 2935Å to 8770Å*, National Bureau of Standards Monograph 61, Washington, D.C., 1966.
4. Sipes, D. L., "Highly efficient neodymium: yttrium aluminum garnet laser end pumped by a semiconductor laser array," *Applied Physics Letters*, Vol. 47, pp. 74-76, July 15, 1985.
5. Atherton, P. D., et al., "Tunable Fabry-Perot filters," *Optical Engineering*, Vol. 20, pp. 806-814, 1981.
6. Russel Austin, R., "The Use of Solid Etalon Devices as Narrow Band Interference Filters." *Optical Engineering*, Vol. 11, pp. 65-69, 1972.
7. Chemical Rubber Company, *Handbook of Chemistry and Physics*, 55th edition, pp. F-166 and F-168.

**Table 1. Doppler shift of solar illumination at epoch of maximum speed away from the sun**

Planet	Maximum speed, km/s	Doppler, pm	Doppler, GHz
Mercury	10.06	18	-18.9
Venus	0.24	0	-0.4
Earth	0.50	1	-0.9
Mars	2.26	4	-4.3
Jupiter	0.63	1	-1.2
Saturn	0.52	1	-1.0
Uranus	0.31	1	-0.6
Neptune	0.04	0	-0.1
Pluto	1.21	2	-2.3
Wavelength (nm):	532.05		
Frequency (THz):	563.5		

**Table 2. Equatorial limb velocity doppler shifts**

Body	Equatorial speed, km/s	Doppler, pm	Doppler, GHz
Sun	2.1	4	-3.9
Earth	0.5	1	-0.9
Mars	0.2	0	-0.5
Jupiter	12.7	22	-23.8
Saturn	10.3	18	-19.4
Uranus	3.9	7	-7.3
Neptune	2.5	4	-4.7
Wavelength (nm):	532.05		
Frequency (THz):	563.5		

**Table 3. Maximum radial component of velocity relative to earth, for circular orbits**

Planet or radius of planetary orbit	Max relative speed, km/s	Doppler, pm	Doppler, GHz
5 solar radii of sun	194.5	345	-365.6
10 solar radii of sun	136.7	243	-256.9
Mercury	36.3	65	-68.3
Venus	13.5	24	-25.4
Earth	0.0	0	.0
Mars	13.9	25	-26.2
Jupiter	27.3	48	-51.3
Saturn	28.8	51	-54.1
Uranus	29.4	52	-55.3
Neptune	29.6	53	-55.6
Pluto	29.7	53	-55.7
Infinity (ecliptic)	29.8	53	-56.0
Wavelength (nm):	532.05		
Frequency (THz):	563.5		

**Table 4. Isosynchronous orbit velocities and doppler shifts**

Body	Altitude, km	Orbital speed, km/s	Doppler, pm	Doppler, GHz
Sun	24055600	73.2	130	-137.6
Mercury	237301	0.3	1	-0.6
Venus	1534720	0.5	1	-0.9
Earth	35775	3.1	5	-5.8
Mars	17032	1.4	3	-2.7
Jupiter	89304	28.2	50	-53.0
Saturn	50964	18.6	33	-35.0
Uranus	36938	9.8	17	-18.4
Neptune	59504	9.1	16	-17.1

Wavelength (nm): 532.05  
Frequency (THz): 563.5

**Table 5. Surface escape velocities and doppler shifts**

Planet or satellite	Escape speed, km/s	Doppler, pm	Doppler, GHz
Mercury	4.2	7	-7.8
Venus	10.4	18	-19.5
Earth	11.2	20	-21.0
Moon	2.4	4	-4.5
Mars	5.0	9	-9.5
Jupiter	60.2	107	-113.2
Io	2.5	4	-4.6
Europa	2.1	4	-3.9
Ganymede	0.9	2	-1.7
Callisto	2.0	4	-3.8
Saturn	36.1	64	-67.8
Titan	0.8	1	-1.5
Iapetus	0.7	1	-1.3
Uranus	22.2	39	-41.7
Titania	0.7	1	-1.4
Neptune	24.5	44	-46.1
Triton	3.1	6	-5.8
Pluto	5.0	9	-9.4

Wavelength (nm): 532.05  
Frequency (THz): 563.5

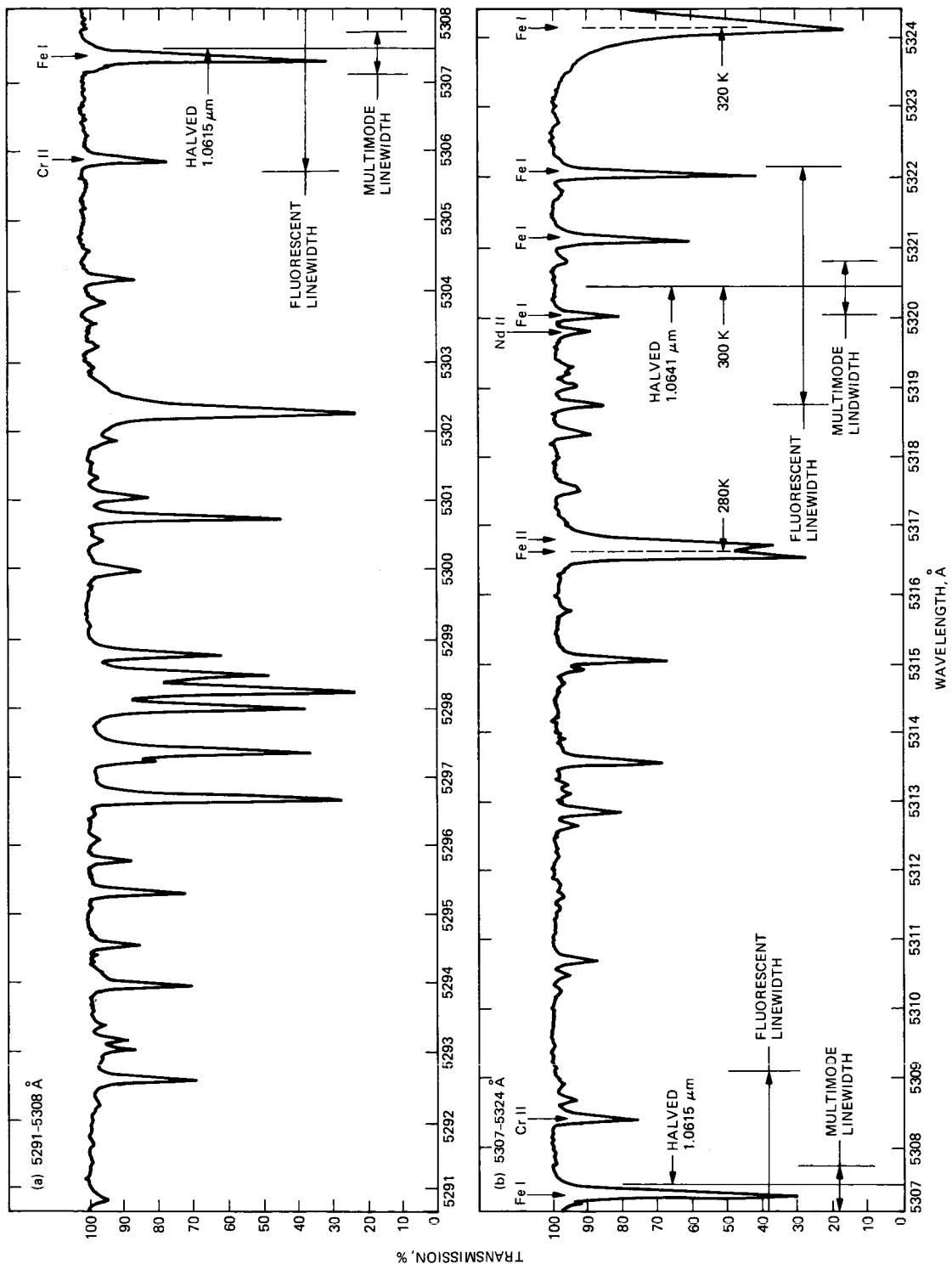


Fig. 1. Solar spectrum from 5291 to 5324 Å

D7-32  
P-9

N87-15336

## A Cost-Performance Model for Ground-Based Optical Communications Receiving Telescopes

J. R. Lesh and D. L. Robinson  
Communications Systems Research Section

*An analytical cost-performance model for a ground-based optical communications receiving telescope is presented. The model considers costs of existing telescopes as a function of diameter and field of view. This, coupled with communication performance as a function of receiver diameter and field of view, yields the appropriate telescope cost versus communication performance curve.*

### I. Introduction

At present, there is much interest in pursuing optical frequencies for deep space communication. Some of the advantages gained through the use of optical communication are: (1) higher data rate communication, (2) smaller size and mass components on the spacecraft compared to equivalent performance radio frequency systems (due to shorter wavelengths), and (3) precise navigational tracking of spacecraft against the stellar background with a single optical receiving station. When considering these advantages, one must also consider the cost of such a system and compare it to the corresponding quantities for other methods of providing equivalent service.

A detailed cost model of an entire deep space optical communication system is very difficult to create. It involves not only the spacecraft and Earth reception ends of the link, but the infrastructures which operate the overall system as well. To build such a model requires a concentrated effort on a number of individual ingredients. This article concentrates on one of those ingredients.

The most desirable location for an optical receiving station is in Earth orbit. However, in all likelihood a ground-based station will both proceed as well as augment an orbiting station. To understand the cost of a ground-based station one must consider also a number of elements. These include the telescope with its mount and pointing control, the dome or protective structure, site preparation, focal plane optics or electronics, control room electronics, ground communications links and the necessary software. Furthermore, recurring costs like maintenance, operations, utilities and spares are also important to consider.

This article considers one of the above elements of a ground-based station: the optical telescope.<sup>1</sup> This element is believed to be one of the major cost ingredients for such a station, and the size and quality of the telescope have a bigger impact on performance than any other station elements.

<sup>1</sup>Telescope costs include costs associated with telescope mount and telescope pointing control.

The general approach to this analysis is as follows: An analytical expression which describes telescope cost as a function of diameter and telescope resolution (determined by the surface quality of the primary and later equated to detector field of view) will first be determined. In order to relate telescope cost to optical communication performance, communication performance is considered as a function of telescope diameter and detector field of view. These two results are coupled together to yield telescope cost as a function of communication performance with telescope diameter and telescope quality as free parameters. From this, the optimum telescope diameter and surface quality are chosen as a function of performance. The result is a final cost versus performance curve for a single ground-based optical telescope. A schematic depicting the approach is given in Fig. 1.

## II. Analysis

In recent years, several large telescope systems have been or are being built. It is from data on cost and performance of these telescopes that projections are based. The diameter and surface quality of the telescope determine the cost of the telescope. In Fig. 2, the costs of various systems have been graphed as a function of diameter for numerous values of surface quality (Ref. 1). In order to facilitate the formulation of an analytical model, straight line approximations have been drawn through points on the graph which correspond to telescopes of similar surface quality. For example, UAM, UKIRT, MMT, SMT and NNTT (see Table 1) all have a blur circle of approximately one arc sec.<sup>2</sup> The line through MMD corresponds to telescopes of 10 arc sec.<sup>2</sup> Finally, the solar collectors and radio telescopes correspond to about 5 arc min.<sup>2</sup> As can be seen from the graph of Fig. 2, similar surface quality telescopes form parallel linear lines on the log/log graph. As the diameter or surface quality increases, telescope cost also increases. An analytical expression which describes these lines can be formulated in the following form:

$$C = \alpha D^x \quad (1)$$

where  $x$  and  $\alpha$  are scaling factors,  $D$  = diameter, and  $C$  = cost.

Values for  $x$  and  $\alpha$  were numerically determined from Fig. 2. It was found that  $x$ , the slope of the parallel linear lines, varied from 2.4 to 2.8, so  $x = 2.6$  was assumed as an average. This is consistent with the RF model which scales as  $x = 2.78$  (Ref. 2).  $\alpha$ , the  $y$  intercept point, was then determined (using  $x = 2.6$ ) to scale as follows:

$$\alpha = F^{-0.94} (10^6)/9.15 \quad (2)$$

---

<sup>2</sup> Aden Meindel and Marjorie Meindel, private communications.

where  $F$  is the telescope resolution or angular blur spot size in arc sec. As an example, consider a 5-m telescope with a 10-arc sec angular blur spot. From Eqs. (1) and (2), the cost of such a telescope is found to be approximately 0.8 million dollars, which agrees reasonably well with the graph in Fig. 2. As a first-order approximation the surface quality given by the blur circle is equated to the field of view of the system limited by the size of the detector, assuming all the light collected is incident on the detector. This is a valid assumption since we are using the telescope as a "light bucket." It is to be noted that the total field of view of the telescope will be much larger. For example, the detector may be optomechanically moved within the total field of view to detect an instantaneous field of view. It is this field of view that is equated to the angular blur circle formed by the telescope and referred to throughout the text.

In order to determine the necessary telescope diameter and field of view, it was necessary to relate these parameters to performance of the resulting communication system. To do this, a reference X-band (8.5 GHz) link<sup>3</sup> was first established, which consisted of a 4.5-m spacecraft antenna with a 10-W X-band transmitter (25 W of raw spacecraft transmitter power) and a 1985 vintage 64-m X-band receiving system. Such a reference system is capable of 26-kbps data transfer from Saturn (range = 10 AU). Then a strawman optical system was defined, which consisted of a 28-cm spacecraft telescope, the same raw spacecraft transmitter power and the same link range. The specific parameters of the reference X-band and strawman optical systems are shown in Appendix A. Using manual and computer analysis tools, the performance of the strawman optical system was then calculated for various values of receiver diameter and field of view. The performance was quantified in terms of achievable data rate. Plots of achievable data rate versus telescope diameter and field of view are shown in Fig. 3. These values were then compared to 26 kbps (the reference system) to determine communications performance gain. A typical achievable data rate example calculation is shown in Appendix B.

Both daylight conditions and moonless night conditions were addressed. However, only the daylight case is shown in Fig. 3. As can be seen from the curves, by combining a larger diameter with a smaller field of view, the system performance is increased. However, as previously observed in Fig. 2, increasing diameter and/or decreasing field of view (increasing surface quality) results in higher cost. Since the highest performance for the least cost is desired, the results of Figs. 2 and 3 must be coupled together and optimized.

---

<sup>3</sup> Layland, J. W., "Conceptual Technology-Capability-Cost Curves," JPL Interoffice Memorandum JWL-85-58, Nov. 11, 1985 (JPL internal document).

Combining Eqs. (1) and (2) with Fig. 3, a third set of curves is obtained which directly relate cost and performance. Cost as a function of performance (measured by increase of data rate capability relative to the reference X-band system) for a 10-m diameter telescope is graphed in Fig. 4. Here, telescope field of view is a parameter and, again for simplicity, only the daylight background case is shown. It can be observed that as higher surface quality (or narrower field of view) is required to obtain a given performance, the cost increases dramatically. A set of curves for 5-m, 10-m, and 15-m telescopes are graphed in Fig. 5. In this figure, both the daylight background and moonless night background cases are included. It can be observed that costs rise sharply as narrower field of view is required. However, it can also be observed that if a specific curve rises sufficiently, using a larger diameter telescope with a wider (poorer surface quality) field of view produces the same performance at a reduced cost. For example, consider 18-dB performance gain during daylight conditions. A 10-m telescope producing such a performance gain would require a surface quality consistent with approximately a 2- $\mu$ rad telescope limited field of view and have a cost of around \$80 million. (Such a telescope would be comparable in size, quality and cost to the Keck telescope which is currently under construction.) However, a 15-m telescope with close to 15 times the surface error (30- $\mu$ rad FOV) could provide equivalent communications performance gain for only \$20 million.

To obtain a single cost versus performance curve, the mean performance gain was selected between daytime and nighttime values for each of the three telescope diameters shown. The cost was then minimized for a given performance gain by optimizing over the telescope diameter. Further smoothing permits interpolation between the discrete values of diameter

treated. Figure 6 summarizes the results of this process. The bracketed line regions of the curve correspond to regions where the identified telescope diameters are expected to apply. Regions between correspond to interpolated diameters. This final cost versus performance curve is estimated to be accurate to within  $\pm 3$  dB.

One issue not examined in this article is that of arraying of smaller telescopes rather than using a single larger telescope. This is because neither the performance models for arraying nor the necessary instrumental costs of arraying are sufficiently well characterized to permit an informative comparison. However, it is believed that arraying will further reduce costs for a given level of performance. This is because arraying costs tend to increase linearly with array area (i.e., as the square of the effective single aperture diameter), whereas the cost model used herein for a single aperture telescope varies as the diameter to the 2.6 power. This issue will be the subject of future investigations.

### III. Concluding Remarks

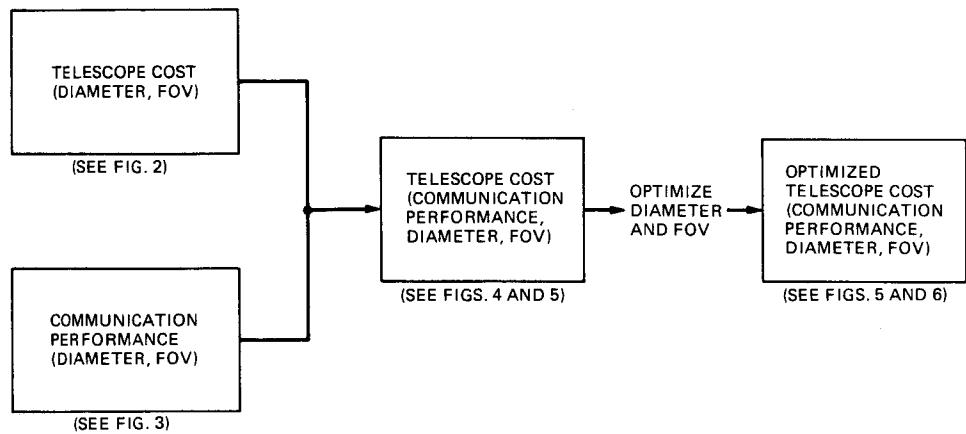
A model has been developed to relate the cost of an optical telescope to the resulting optical communications performance. This was done by relating available telescope cost data to a pair of parameters, the telescope diameter and a measure of telescope quality. These same quantities were used in a communications model context to quantize the communications performance gain relative to an RF reference system. The cost and performance were then related through equivalent values of telescope diameter and quality. The final cost model then resulted by picking the telescope diameter (and associated telescope quality) which minimized telescope cost for a given amount of performance improvement.

## References

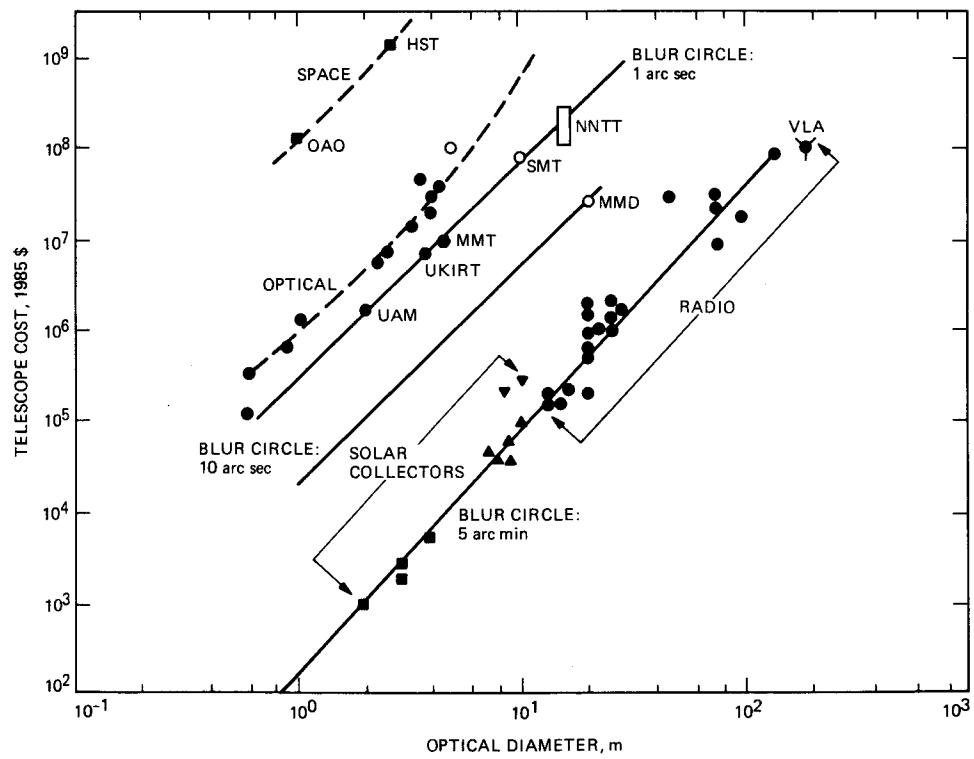
1. Meinel, Aden and Meinel, Marjorie, "Very Large Optics of the Future," *Optics News*, pp. 9-14, Mar. 1986.
2. Potter, P. D., Merrick, W., Ludwig, A., *Large Antenna Apertures and Arrays for Deep Space Communications*, Technical Report 32-848, Jet Propulsion Laboratory, Pasadena, CA, Nov. 1, 1965.
3. Lambert, S. A., et al., *Design and Analysis Study of a Spacecraft Optical Transceiver Package*, McDonnell Douglas Astronautics Co., Final Report under JPL contract No. 967061, August 19, 1985.

**Table 1. Acronyms used in text and figures**

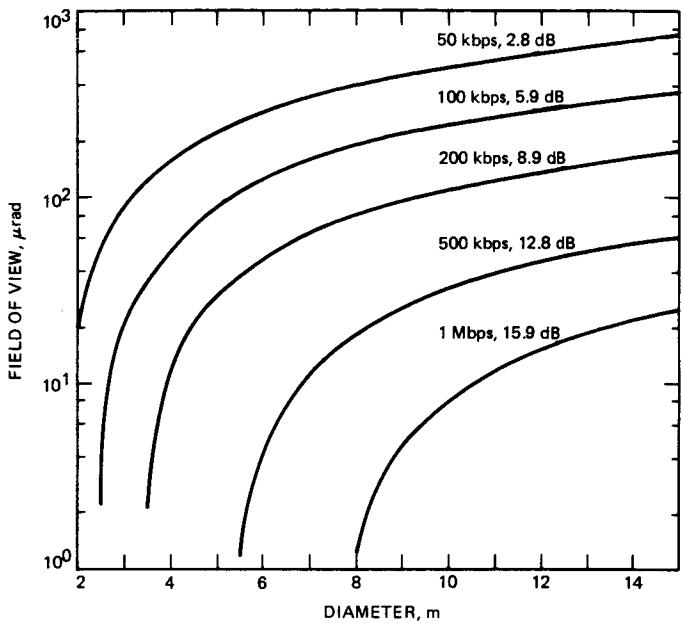
HST	Hubble Space Telescope
MMD	Millimeter Dish
MMT	Multimirror Telescope
NNTT	National New Technology Telescope
OAO	Orbiting Astronomical Observatory (Copernicus)
OPTRANSPAC	Optical Transceiver Package
SMT	Sub-Millimeter Telescope
UAM	University of Mexico
UKIRT	United Kingdom Infrared Telescope
VLA	Very Large Array



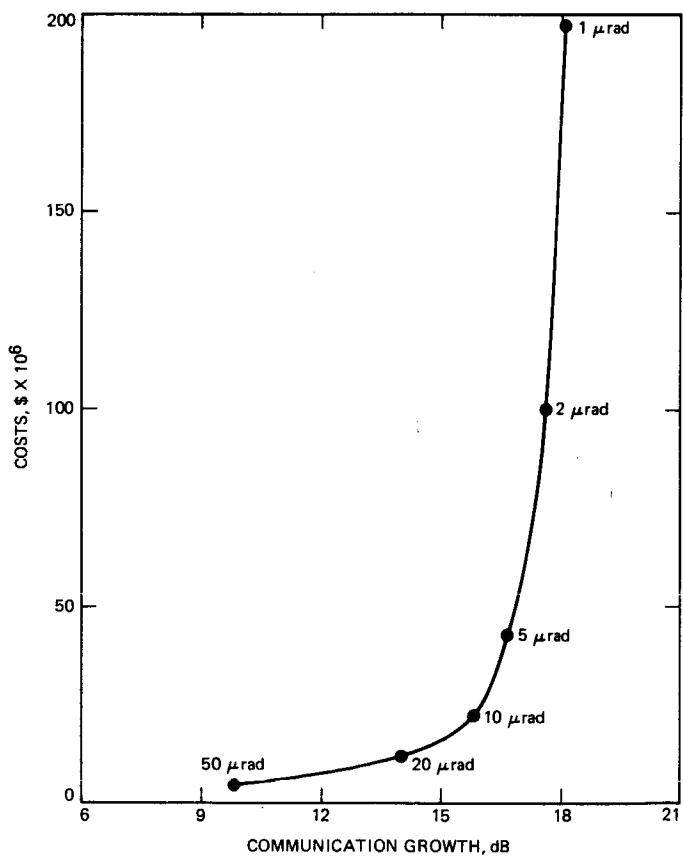
**Fig. 1. Outline of approach for cost-performance modeling of optical communications receiving telescope**



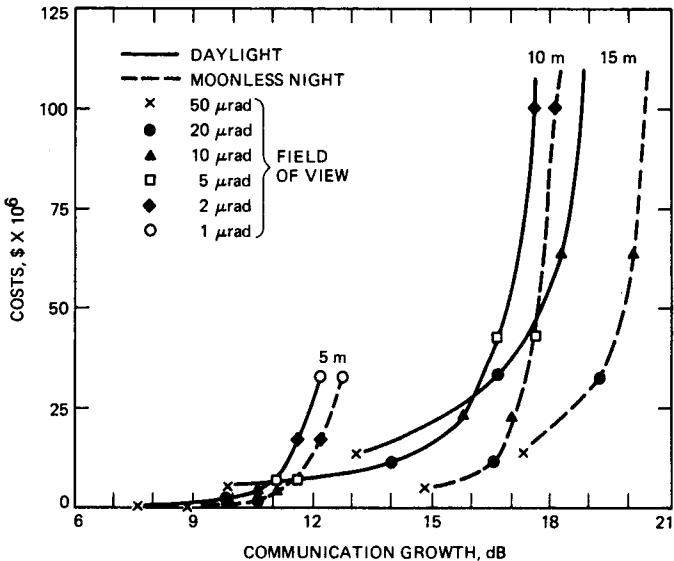
**Fig. 2. Telescope costs as a function of diameter and angular blur circle resolution (see Table 1 for definition of acronyms)**



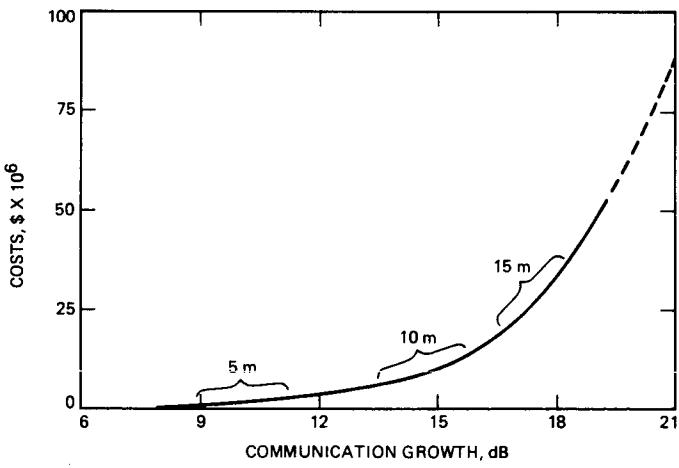
**Fig. 3.** Achievable optical communication data rates as a function of telescope diameter and field of view for daylight conditions (see Appendix A for specific operational characteristics)



**Fig. 4.** Costs of 10-m telescope as a function of field of view and performance for daylight viewing conditions



**Fig. 5.** Costs vs performance of 5-, 10-, and 15-m telescopes for both daylight and moonless night conditions



**Fig. 6.** Optimized telescope cost curve

## Appendix A

### **Reference X-band (8.5 GHz) and Strawman Optical System Description**

This appendix describes both the reference X-band (8.5 GHz) system and the static portions of the strawman optical system used in this report.

#### **I. Reference X-band System**

The reference system against which optical communications performance was compared uses a 4.5-m spacecraft antenna and a radio frequency wavelength of 3.5 cm (X-band). The spacecraft transmitter produces 10 W of X-band output power with a power conversion efficiency of 40%. This gives rise to a 25-W overall power consumption of the transmitter. The communications range was from Saturn (10 AU), and the reception system was the 1985 version of a 64-m DSN station. This overall system is capable of achieving 26 kbps from Saturn (see Footnote 3).

#### **II. Strawman Optical System**

The optical link considered consists of a 27.9-cm telescope with a frequency doubled Nd:YAG laser at the spacecraft. The transmitted wavelength is 0.532  $\mu\text{m}$ . The same raw transmitter power as the reference systems was assumed (25 W). This, coupled with an assumed laser power conversion efficiency, yields 2 W of laser power. The spacecraft was assumed to be at Saturn (10 AU range) with the planet directly behind the spacecraft. The values for transmitter diameter, transmitter obscuration (0.0615 m), transmitter pointing bias error (0.3  $\mu\text{rad}$ ), and transmitter rms pointing jitter (0.233  $\mu\text{rad}$ ) were taken from the OPTRANSPAC (Optical Transceiver Package) study (Ref. 3) performed by McDonnell Douglas. Atmospheric transmission, transmitter optics efficiency, receiver optics efficiency, and the quantum efficiency of the detector (with a spectral bandwidth of 10 angstroms) were each 50%. The receiver diameter, field of view and data rate were variable parameters to be evaluated. Pulse position modulation with  $M = 256$ , and a slot width of 10 ns were specified. The link performance was specified at an uncoded bit error rate of  $10^{-3}$ , a value which can be made much smaller than  $10^{-6}$  with even a moderate amount of coding.

## Appendix B

### Sample Calculation of Optical Link Performance

Antennae parameters:

Wavelength, $\mu\text{m}$	=	0.532
Transmitted average power, W	=	2.00
Distance between XMTR and RCVR, AU	=	10.0
Amospheric transmission fraction	=	0.500
Diameter of XMTR, m	=	0.279
Obscuration diameter of XMTR, m	=	$0.615 \times 10^{-1}$
XMTR optics efficiency	=	0.500
XMTR pointing bias error, $\mu\text{rad}$	=	0.300
XMTR rms pointing jitter, $\mu\text{rad}$	=	0.233
Diameter of RCVR, m	=	10.0
Obscuration diameter of RCVR, m	=	0.0
RCVR optics efficiency	=	0.500
Narrow band filter transmission	=	1.00
Spectral bandwidth, $\text{\AA}$	=	10.0
RCVR diam, field of view, $\mu\text{rad}$	=	10.0

PPM detector parameters:

Alphabet size ( $M = ?$ )	=	256.0
Data rate, kbits/s	=	26.0
Dead time, $\mu\text{s}$	=	305.0
Slot width, ns	=	10.0
Desired probability of bit error	=	$1.0 \times 10^{-3}$
Quantum efficiency of detector	=	0.500

Noise sources:

$$\text{Saturn} - \text{RCVR to source distance, AU} = 10.0$$

Additional noise sources:

$$\text{DAYLIGHT} - \text{radiance, W/m}^2/\text{sr}/\text{\AA} = 0.323 \times 10^{-2}$$

**LINKOUTPUT**

		<u>Factor</u>	<u>dB</u>
Transmitted power, W		2.00	33.0 dBm
Min Reqd peak power, W	= $0.62 \times 10^5$		
XMTR antenna gain		$0.187 \times 10^{13}$	122.7
XMTR diam, m	= 0.279		
Obscuration diam, m	= 0.062		
Beam width, $\mu$ rad	= 3.328		
XMTR optics efficiency		0.500	-3.0
XMTR pointing efficiency		0.878	-0.6
Pointing error, $\mu$ rad	= 0.300		
RMS jitter, $\mu$ rad	= 0.233		
Space loss (10.00 AU)		$0.801 \times 10^{-39}$	-391.0
Atmospheric transmission		0.500	-3.0
RCVR antenna gain		$0.349 \times 10^{16}$	155.4
RCVR diam, m	= 10.000		
Obscuration diam, m	= 0.000		
Field of view, $\mu$ rad	= 10.000		
RCVR optics efficiency		0.500	-3.0
Narrow band filter transmission		1.00	0.0
Bandwidth, Å	= 10.000		
Received signal power, W		$0.114 \times 10^{-11}$	-89.4 dBm
Received noise power, W	= $0.548 \times 10^{-10}$		
Quantum efficiency		0.500	-3.0
Photons/joule		$0.268 \times 10^{19}$	184.3 dB/J
Received signal photoelectrons (PE)/s		$0.153 \times 10^7$	61.9 dB-Hz
Symbol time, s		$0.308 \times 10^{-3}$	-35.1 dB-Hz
Received signal PE/symbol		472.0	26.7
Required signal PE/symbol		13.7	11.4
Received noise PE/slot	= 0.735		
Margin		34.6	15.3

# Relative Performance of 8.5-GHz and 32-GHz Telemetry Links on the Basis of Total Data Return per Pass

M. A. Koerner

Telecommunications Systems Section

*The performance of X-band (8.5-GHz) and 32-GHz telemetry links is compared on the basis of the total data return per DSN station pass. Differences in spacecraft transmitter efficiency, transmit circuit loss, and transmitting antenna area efficiency and pointing loss are not considered in these calculations. Thus, the performance differentials calculated in this memo are those produced by a DSN 70-m station antenna gain and clear weather receiving system noise temperature and by weather.*

*These calculations show that, assuming mechanical compensation of the DSN 70-m antenna for 32-GHz operation, a performance advantage for 32 GHz over X-band of 8.2 dB can be achieved for at least one DSN station location. Even if only Canberra and Madrid are used, a performance advantage of 7.7 dB can be obtained for at least one DSN station location. A system using a multiple beam feed (electronic compensation) should achieve similar results.*

## I. Introduction

The various contributions, positive and negative, to the performance differential between X-band and 32-GHz telemetry links for interplanetary missions can be divided into spacecraft-related contributions and DSN-related contributions. The spacecraft-related contributions are the differences in transmitter efficiency, transmitting circuit loss, and the transmitting antenna area efficiency and pointing loss. These differences are very dependent on the spacecraft mission and the hardware to be employed. For example, the difference between X-band and 32-GHz TWTA efficiency may be significantly less than the difference between X-band and 32-GHz solid-state power amplifier efficiency, and the differential in spacecraft antenna pointing loss may be a strong function

of the antenna diameter. These differences also may decrease significantly as technology improves. The DSN-related contributions are those arising from the difference in DSN station antenna gain and clear-weather system noise temperature, weather effects, and DSN antenna pointing loss. These differences are mission independent and should change less as improved technology becomes available than the spacecraft-related contributions.

The objective of the calculations in this report is to establish the net DSN-related contributions to the performance differential between X-band and 32-GHz telemetry links. As a further simplification, the differential in DSN antenna pointing loss used in all these calculations assumes the X-band and 32-GHz DSN 70-m station antenna pointing errors are 0.003°

and  $0.001^\circ$ , producing X-band and 32-GHz DSN antenna pointing losses of  $-0.1$  dB and  $-0.17$  dB, respectively. Thus the difference between 32-GHz and X-band DSN 70-m station antenna pointing loss is only  $-0.07$  dB for all of the results shown in this report. Thus, these calculations provide a baseline to which the spacecraft-related performance differentials for transmitter efficiency, transmitting system circuit loss, and transmitting antenna area efficiency and pointing loss, as well as the difference in DSN 70-m station antenna pointing loss, must be added.

A major choice to be made in calculating the DSN-related contributions to the 32-GHz to X-band performance differential is that of the parameter used to measure link performance. Typically, achievable data rate has been used. However, as the DSN 70-m station antenna gain and clear-weather system noise temperature and the weather effects all depend on the DSN station elevation angle, the achievable data rate for both X-band and 32-GHz links will be a function of elevation angle. The ratio of 32-GHz to X-band achievable data rate could be computed as a function of elevation angle, but one is left with a somewhat arbitrary choice of elevation angle. Typically, a  $30^\circ$  elevation angle has been used.

In this report the number of bits returned (total data return) per DSN station pass is used as the measure of link performance. Thus, the ratio of 32 GHz to X-band link performance is the ratio of the corresponding total data returns per DSN station pass. This ratio is computed for the total data returns obtained using (1) the best fixed data rate, (2) the best two data rates, and (3) a continuously variable data rate. The best fixed data rate is the single data rate which yields the greatest total data return per pass. The total data return for the best two rates is the result of a similar calculation, when two data rates, with one increase and one decrease in data rate per pass, can be used. The continuously variable data rate provides an upper bound in performance. This system continuously uses the maximum rate allowed by the instantaneous performance.

The calculation of the 32-GHz to X-band performance advantage for these three different levels of operational complexity gives some insight into the effect of data rate strategy on the ratio of 32-GHz to X-band link performance. As we shall see, because of the greater sensitivity of 32-GHz links to elevation angle, the 32-GHz to X-band performance advantage increases with the number of data rates one can employ during a DSN station pass.

As the spacecraft declination (with respect to Earth) and the DSN station location determine the elevation angle profiles (elevation angle versus time) for the station and 32-GHz links are affected more by elevation angle than X-band links, the

ratios of 32-GHz to X-band performance calculated in this article vary with declination and DSN station location. As we shall see, the ratio of 32-GHz to X-band link performance increases with increasing declination for Goldstone and Madrid, the northern hemisphere stations, and decreases with increasing declination for Canberra, the southern hemisphere station.

## II. DSN Antenna Gain Models

The five DSN 70-m station antenna gain versus elevation angle models used in these calculations are shown in Fig. 1. There is one X-band model and four 32-GHz models. The DSN 70-m station antenna gains shown in Fig. 1 include the clear-weather atmospheric attenuation.

### A. X-Band Model

The X-band DSN 70-m station antenna gain versus elevation angle model shown in Fig. 1 is that specified by the DSN for use by the VRM project, less  $0.1$  dB to allow for diplexing loss. This model is also being used for the MM II/CRAF link performance calculations.

### B. 32-GHz/Baseline Model

This 32-GHz antenna gain model is an estimate of the 32-GHz performance of a DSN 70-m station with no improvements for 32 GHz. This model was obtained by a somewhat different process than the X-band model discussed above and may represent a somewhat more optimistic view of the unimproved DSN 70-m station performance.

For the 32-GHz/Baseline antenna gain model, the net antenna gain will be

$$G_R = G_{RB} + L_{FX} + L_{GV} + L_{TU} + L_{ATM} \quad (1)$$

where  $G_{RB}$  is  $87.412$  dB, the gain at 32 GHz of a 70-m parabolic antenna with 100% area efficiency,  $L_{FX}$  is the sum of the fixed losses, those that do not vary with elevation angle,  $L_{GV}$  is the loss due to gravitational deformations,  $L_{TU}$  is the loss due to atmospheric turbulence, and  $L_{ATM}$  is the clear-weather atmospheric attenuation. For the 32-GHz/Baseline antenna gain model,  $L_{FX}$  is  $-3.841$  dB. The factors contributing to this fixed loss are listed in Table 1.

The loss  $L_{GV}$  at 32 GHz from gravitational deformation of the DSN 70-m antenna surface is extrapolated from estimates for X-band. The assumption is that

$$L_{GV} = 10 \log_{10} [\exp(-(4\pi\sigma_{GV}/\lambda)^2)] \quad (2)$$

where  $\sigma_{GV}$  is the standard deviation of the surface deformations caused by gravity and  $\lambda$  is the RF wavelength. Given this assumption,

$$(L_{GV}) \text{ 32 GHz} = 14.4608 (L_{GV}) \text{ X-band} \quad (3)$$

where 14.4608 is the square of the ratio of 32 GHz to the X-band RF frequency (8.415 GHz). Using this approach, one obtains the data in Table 2. Values of  $L_{GV}$  for values of elevation angle between those given in Table 2 are computed using second-order (quadratic) interpolation. The resulting values of  $L_{GV}$  are plotted as a function of elevation angle in Fig. 2.

In these calculations, the DSN 70-m antenna gain reduction from atmospheric turbulence  $L_{TU}$  is calculated following the approach used in Ref. 1. Sufficient data is given in Ref. 1 for this loss to be computed at 10°, 30°, and 90° elevation angles. At 10° elevation angle,  $L_{TU}$  is -0.878 dB. At 30° elevation angle,  $L_{TU}$  is -0.524 dB. At 90° elevation angle,  $L_{TU}$  is -0.142 dB. Second-order (quadratic) interpolation is used to compute  $L_{TU}$  for values of elevation angle other than 10°, 30°, and 90°. The resulting values of  $L_{TU}$  are plotted as a function of elevation angle in Fig. 2.

As noted above, the DSN 70-m antenna gains shown in Fig. 1 include the loss  $L_{ATM}$  from the clear-weather atmospheric attenuation. For all four 32-GHz models,

$$L_{ATM} = -0.081/\sin(\text{ELE}) \text{ dB} \quad (4)$$

where -0.081 dB is the average of the clear-weather atmospheric attenuations for Goldstone (-0.079 dB) and the overseas stations (-0.083 dB) obtained from the S. Slobin 32-GHz weather model (see Subsection 4). The resulting values of  $L_{ATM}$  are plotted as a function of elevation angle in Fig. 2.

### C. 32-GHz/Passive Improvements Model

The DSN 70-m station 32-GHz/Passive Improvements antenna gain model differs from the DSN 70-m station 32-GHz/Baseline antenna gain model only in the reduction of  $L_{FX}$ , the sum of those losses that do not vary with elevation angle, by 1.43 dB from -3.841 dB to -2.411 dB.

This 1.43-dB improvement is comprised of a 0.1-dB reduction in quadrapod blockage, 0.2 dB from stiffening of the antenna structure to resist deflections caused by wind, 0.81 dB from more accurate setting (0.203 mm (0.008 in.) rms) of the panels which make up the main reflector surface, and 0.32 dB from the use of a new, more accurate (0.152 mm (0.006 in.) rms) subreflector.

### D. 32-GHz/Mechanical Compensation Model

With active mechanical compensation for deflections of the DSN 70-m antenna surface, the loss due to gravitational deflections of the antenna surface can be reduced to -0.126 dB, independent of elevation angle, and the losses from wind and thermal distortions can be reduced by 0.5 dB and 0.4 dB, respectively. These improvements together with the 0.1-dB reduction in quadrapod blockage, 0.81 dB from more accurate setting of the main reflector panels, and 0.32 dB from a more accurate subreflector, yield an antenna gain of

$$G_R = G_{RB} + L_{FX} + L_{TU} + L_{ATM} \quad (5)$$

where in this case  $L_{FX}$  is -1.837 dB and  $G_{RB}$ ,  $L_{TU}$ , and  $L_{ATM}$  are the same as for the DSN 70-m station 32-GHz/Baseline antenna gain model discussed above.

### E. 32-GHz/Electronic Compensation Model

With the use of a multiple-beam, cryogenically-cooled feed, the potential exists for reduction of the losses from atmospheric turbulence as well as those from gravitational, wind, and thermal distortion of the main reflector surface. The calculations presented in this report assume that, temporarily neglecting atmospheric attenuation, the antenna area efficiency  $L_{EC}$  is -2.22 dB (60%) at a 45° elevation angle and -3.01 dB (50%) at elevation angles of 10° and 90°. Second-order (quadratic) interpolation is used to obtain  $L_{EC}$  at elevation angles other than 10°, 45°, and 90°. Having calculated  $L_{EC}$ , the net DSN 70-m station antenna gain for the 32-GHz/Electronic Compensation Model is

$$G_R = G_{RB} + L_{EC} + L_{ATM} \quad (6)$$

where  $G_{RB}$  and  $L_{ATM}$  are as discussed above for the DSN 70-m station 32-GHz/Baseline antenna gain model.

## III. Clear-Weather System Noise Temperature Models

The DSN 70-m station X-band and 32-GHz clear-weather receiving system noise temperature models used in these calculations are based on the DSN 64-m station X-band clear-weather receiving system noise temperature model.<sup>1</sup> This reference specifies that the noise temperature for a non-diplexed (listen-only) system is 20 K at zenith (90° elevation angle) and the increase above the zenith noise temperature, for

<sup>1</sup>Deep Space Network/Flight Project Interface Design Handbook, JPL Document 810-5: Vol. II, Module TCI-10, Sept. 1, 1981, Jet Propulsion Laboratory, Pasadena, Calif. (JPL internal document).

elevation angles other than 90°, is that given by the "X-band" curve in Fig. 3.

### A. X-Band

The DSN 70-m station X-band clear-weather receiving system noise temperature model used in these calculations is for diplexed operation. The diplexer is expected to increase the noise temperature about 5 K. Thus, these calculations assume that the DSN 70-m station X-band clear-weather receiving system noise temperature at zenith is 25 K and that the noise temperature increase above the zenith noise temperature, for elevation angles other than 90°, is the same as that for the DSN 64-m nondiplexed station, shown as the curve labeled "X-band" in Fig. 3.

### B. 32 GHz

With the exception of the atmospheric contribution, these calculations assume the DSN 70-m station 32-GHz clear-weather receiving system noise temperature is the same as the DSN 64-m station clear-weather receiving system noise temperature. After correction for the difference in the X-band and 32-GHz atmospheric contributions, the DSN 70-m station 32-GHz clear-weather receiving system noise temperature at zenith is 23.15 K, and the increase above the zenith noise temperature, for elevation angles other than 90°, is given by the curve labeled 32 GHz in Fig. 3.

## IV. Weather Degradation Model

A weather model provides a means of calculating the cumulative probability distribution of the weather degradation for different DSN station locations and elevation angles. The weather degradation in decibels is the sum of the incremental atmospheric attenuation in decibels, above that for clear weather, and the ratio in decibels of the system noise temperature with weather of a given cumulative probability to the clear-weather system noise temperature. Results are presented in this article for four different weather models. All of the models were created by S. D. Slobin of JPL's Radio Frequency and Microwave Subsystems Section. One of the weather models, hereafter referred to as the Slobin/810-5 weather model,<sup>2</sup> was developed as a X-band weather model and is currently being used for the MM II/CRAF X-band link performance calculations.

The other three models, hereafter designated the Slobin/Best, Slobin/Average, and Slobin/Worst weather models, are based on a combination of K-band radiometer measurements at Goldstone at 30° elevation angle and cloud-cover and rainfall statistics for sites similar to the DSN station locations. The Slobin/Average weather model is used to calculate most of the results presented in this article. For these three models,

the Canberra and Madrid weather statistics are the same. Table 3 tabulates the 32-GHz noise temperature increase at 30° elevation angle as a function of cumulative probability for the Slobin/Best, Slobin/Average, and Slobin/Worst weather models for both Goldstone and the two overseas sites. Using the data in Table 3 for the selected model as a starting point, the Slobin/810-5 weather model methodology<sup>2</sup> can be employed to calculate the weather degradation as a function of cumulative probability for any desired elevation angle and DSN station location.

As the objective of this article is to compare X-band and 32-GHz link performance, it is desirable to use the same weather model for both the X-band and 32-GHz link calculations. If one assumes, as this article does, that the weather effects are entirely caused by water droplets in clouds and by rain, the 32-GHz atmospheric attenuation due to weather will be 14.4608 times the X-band atmospheric attenuation. The factor 14.4608 is the square of the ratio of the 32-GHz and X-band (8.415-GHz) RF frequencies. Using this relationship between the X-band and 32-GHz atmospheric attenuations caused by weather, one can easily compute the 32-GHz weather degradation from the X-band weather degradation or vice versa. However, because of the large multiplication factor, a small error in a X-band weather degradation can create a very large error in the 32-GHz weather degradation. Coupled with the very qualitative observation that the X-band Slobin/810-5 model appears slightly conservative, this suggests that one — Slobin/Best, Slobin/Average, or Slobin Worst — of the weather models based partly on 32-GHz radiometer measurements at Goldstone is a more appropriate weather model for the comparisons of X-band and 32-GHz link performance than the Slobin/810-5 weather model. As noted previously, the Slobin/Average weather model is used for the bulk of the calculations presented in this report.

## V. Calculation of Achievable Data Rate

Having defined the models to be used for the DSN 70-m station X-band and 32-GHz antenna gain and clear-weather receiving system noise temperature and models to be used to calculate the weather degradation, the next step is to use these models to calculate achievable data rate. Sample calculations of achievable data rate for X-band and 32-GHz links and the Canberra DSN 70-m station are shown in Tables 4 and 5. These sample calculations are for a 30° elevation angle (listed under item 10 in each table). The achievable data rate is listed under item 13 in each table.

<sup>2</sup>Deep Space Network/Flight Project Interface Design Handbook, JPL Document 810-5: Vol. I, Module TCI-40, Rev. B, Dec. 1, 1983, Jet Propulsion Laboratory, Pasadena, Calif. (JPL internal document).

Items 1 through 9 of each table list the RF link parameter values used in this calculation and item 10 shows the available ratio of total received power to receiving system noise spectral density ( $P_T/N_O$ ) for clear weather. Note that, as discussed in the introduction, the transmitting system RF power output, circuit losses, and antenna pointing losses in Tables 4 and 5 are the same. The transmitting antenna gains in Tables 4 and 5 differ only by the square of the ratio of 32-GHz to the X-band link RF frequency (8.415 GHz). Thus, the transmitting antenna area efficiencies are the same. Furthermore, as discussed in the introduction, the DSN antenna pointing loss in the X-band link performance estimate in Table 4 is -0.10 dB (for 0.003-degree pointing error) and that for the 32-GHz link performance estimate in Table 5 is -0.17 dB (for 0.001-degree pointing error). Thus, the ratio of 32-GHz to X-band DSN antenna pointing loss in Tables 4 and 5 and in all the other numerical results shown in this article is -0.07 dB.

In the link performance calculations shown in Tables 4 and 5, the mean of the performance margin (item 14 in Tables 4 and 5) has been adjusted to make the resulting link reliability (item 15 in Tables 4 and 5) equal to 0.95. The link reliability is the probability that the mean, clear-weather performance margin is greater than the deviation from mean, clear-weather link performance due to both weather and link parameter variations. Then, as there is no ranging suppression, the mean required  $P_T/N_O$  (item 13 in Tables 4 and 5) can be, at most, the difference between the mean available  $P_T/N_O$  (item 10) and the mean, clear-weather performance margin (item 14). Given the RF receiver threshold noise bandwidth shown near the top of Tables 4 and 5 and the required carrier margin and  $E_B/N_O$ , listed under item 13 in Tables 4 and 5, the achievable data rate is determined.

The computation of the link reliability is a matter of finding that value of the cumulative probability distribution of the sum of the degradation from weather and link parameter variations that corresponds to a degradation equal to the mean clear-weather performance margin. In this case the inverse calculation is needed. One starts with the required link reliability, and needs to calculate the sum of the potential link degradations from weather and link parameter variations that has that cumulative probability.

A piecewise linear approximation is used for the cumulative probability distribution of the weather degradation, with the break-points calculated using the selected weather model. For the Slobin/810-5 weather model, the cumulative probabilities at which the break-points (discontinuities in slope) occur are those given in Table 1 of footnote 2. For the Slobin/Best, Slobin/Average, and Slobin/Worst weather models, the cumulative probabilities at which the breakpoints occur are those in Table 3 of this article (11 breakpoints, not

including zero probability). The variance of a parameter's variation is computed for each of the link parameters, assuming the parameter either has a uniform distribution or a triangular distribution between the limits defined by the positive and negative tolerances shown in Tables 4 and 5. For the triangular distribution, the peak of the triangle is at the parameter design value. The distribution used is designated by a "U" (uniform) or "T" (triangular) in the "DIST" column of Tables 4 and 5. The resulting variances, shown under the "variance" column in Tables 4 and 5, are added to yield the variance for the clear-weather performance margin.

The corresponding standard deviation "SIGMA" is listed under the link reliability (Item 15) in the "mean" column. Note that the "SIGMA" in Table 5 is the same as that in Table 4. The 32-GHz link tolerances in Table 5 are typical of those used at the time of a project start and do not reflect current uncertainties. Using the assumption that the sum of the parameter variations is Gaussian (central limit theorem) and the piecewise approximation for the cumulative probability distribution of the weather degradation, an expression has been derived for the desired cumulative probability distribution. The resulting expression is a summation which requires the evaluation of one exponential function and one error function per breakpoint in the piecewise linear approximation.

By repeating the calculation shown in Tables 4 and 5 for different elevation angles and calculating the ratio of 32-GHz to X-band achievable data rate, one can produce curves such as those shown in Fig. 4. In Fig. 4, the ratio of 32-GHz to X-band achievable data rate is plotted as a function of elevation angle for Goldstone, Canberra, and Madrid for link reliabilities of 0.90 and 0.95. The results shown in Fig. 4 are for the DSN 70-m station 32-GHz/Mechanical Compensation antenna gain model, diplexed X-band, and Slobin/Average weather. With such a comparison, however, the problem remains as to which elevation angle is "significant," since during a DSN station pass the elevation angle may vary over nearly the full range shown in Fig. 4.

## VI. Calculation of Total Data Return Per Pass

The first step in calculating total data return per pass is to calculate achievable data rate as a function of time during the pass. For a given declination, the DSN station elevation angle can be computed as a function of time for the three DSN station locations. Combining such a calculation with a calculation of achievable data rate, similar to those shown in Tables 4 and 5, yields achievable data rate as a function of time during a one-day period for the three DSN station locations. The result of a sample calculation for 0° declination is shown in Fig. 5. These curves were computed using 51 points, evenly

spaced in time, starting with the time of minimum elevation angle, which is  $10^\circ$ , and ending with the time of peak elevation angle. The curve for a DSN station location is symmetric about the time of peak elevation angle.

Results are shown in Fig. 5 for X-band and 32-GHz links, 0.90 and 0.95 link reliability, and the three DSN station locations. With the exception of those link parameters dependent on elevation angle and/or link reliability, the link parameter values for the X-band results in Fig. 5 are those shown in Table 4 and the link parameter values for the 32-GHz results in Fig. 5 are those shown in Table 5. The 32-GHz/Mechanical Compensation DSN 70-m station antenna gain model and the Slobin/Average weather model were used. The range for the performance estimates in both Tables 4 and 5 and Fig. 5 was 10 AU.

Obtaining the total data return from achievable data-rate profiles for a one-day period requires the selection of a data-rate strategy. The selection of the data-rate strategy depends on the amount of operational complexity permitted. In this article, 32-GHz and X-band performance will be compared for three different data-rate strategies.

The fixed-rate strategy allows one to use any data rate during a DSN station pass, but the rate must remain fixed during the pass. For the fixed-rate strategy, the total data return is the product of the selected data rate and the time per pass this data rate can be supported with the required link reliability. The results shown in this article assume that the best fixed rate is selected. Thus, the comparisons of X-band and 32-GHz total data return for the fixed-rate strategy are made using the best fixed-rate total data returns per pass.

The two-rate strategy allows the use of two data rates per pass, with one increase and one decrease in data rate per pass. As in the fixed-rate strategy, these calculations assume that the best two rates would be used. Thus, the comparisons of 32-GHz and X-band total data return for the two-rate strategy are made on the basis of the best two-rate total data returns per pass.

The variable-rate strategy allows a continuously variable data rate during a DSN station pass. For this strategy, the total data return per pass is simply the integral of the achievable data rate versus time. The integral is calculated using a trapezoidal approximation with 101 points per pass. While this strategy would never be used with the current DSN telemetry hardware, comparisons of X-band and 32-GHz total data return per pass for a variable data-rate strategy do provide an upper bound on the increase in the 32-GHz to X-band performance advantage to be achieved by using more than two data rates per DSN station pass.

## VII. Numerical Results

With the exception of those parameters dependent on elevation angle and/or link reliability, the X-band and 32-GHz link parameter values used to calculate the results shown in this section are the same as those shown in Tables 4 and 5. However, since the ratio of 32-GHz to X-band total data return is being computed, the absolute values of the X-band and 32-GHz spacecraft-related link parameters used for these calculations are not important, as long as their relative values remain the same.

To simplify figure labeling, R1, R2, and RV will be used for the ratio of the 32-GHz to X-band total data return per pass using the best fixed (one) data rate, the best two data rates, or a variable data rate during a DSN station pass. Note that all values of R1, R2, and RV shown in this article are expressed in decibels.

### A. Comparison of R1, R2, and RV

Figure 6 shows curves of R1, R2, and RV as a function of spacecraft declination for the three DSN station locations. These results assume a mechanically compensated DSN 70-m antenna for 32 GHz, the DSN 70-m antenna is diplexed at X-band, the Slobin/Average weather model, and a 0.95 required link reliability. The primary purpose of this figure is to show, at least for a mechanically compensated DSN 70-m antenna at 32 GHz, that the difference between the performance advantages of 32 GHz over X-band for the fixed-rate and variable-rate strategies is almost independent of both declination and DSN station location and is approximately 1 dB. The difference between the performance advantages of 32 GHz over X-band for the fixed-rate and two-rate strategies is also almost independent of both declination and DSN station location and appears to be about 0.2 dB. The rest of the results shown in this report consider only R1, the 32-GHz over X-band performance advantage for a fixed- (one) rate data-rate strategy. However, it is important to remember, in examining subsequent figures, that a more complex data-rate strategy could improve the performance advantage of 32 GHz over X-band from 0.2 dB to 1.0 dB.

### B. Effect of Link Reliability

Figure 7 shows curves of R1 as a function of declination for the three DSN station locations and link reliabilities of 0.90 and 0.95. These results assume a mechanically compensated DSN 70-m antenna for 32 GHz, diplexed operation at X-band, and the Slobin/Average weather model. Clearly, decreasing the required link reliability increases the performance advantage of 32 GHz over X-band. The difference be-

tween the performance advantage of 32 GHz over X-band for 0.90 link reliability and that for 0.95 link reliability varies somewhat with declination. For Goldstone the difference is 0.39 dB at  $-25^\circ$  declination and 0.14 dB at  $+25^\circ$  declination. For Canberra the difference is 0.26 dB at  $-25^\circ$  declination and 0.61 dB at  $+25^\circ$  declination. For Madrid the difference is 0.73 dB for  $-25^\circ$  declination and 0.26 dB at  $+25^\circ$  declination. Note that the difference for Goldstone is significantly less than that for Canberra and Madrid, and that, for all three DSN station locations, the largest of these differences occurs for the least favorable declination ( $-25^\circ$  declination for Goldstone and Madrid, which are northern hemisphere stations, and  $+25^\circ$  for Canberra, which is a southern hemisphere station).

The reason for these differences is that the allowance that must be made for weather and link parameter variations is much greater for 32-GHz links than for X-band links. Examining item 14 in Tables 4 and 5, the sample X-band and 32-GHz link performance estimates, this allowance is 1.38 dB for the X-band link performance estimate in Table 4 and 4.27 dB for the 32-GHz link performance estimate in Table 5. As the allocation for weather and link parameter variations will be much greater at 32 GHz than at X-band for any required link reliability, one would logically expect that the difference between the allocations for two different link reliability levels will be much greater at 32 GHz than at X-band. The difference, for 0.90 and 0.95 link reliability, between the 32-GHz difference in allocation and the X-band difference in allocation is the separation of the curves in Fig. 7. The reason the separation between curves in Fig. 7 (for a given station location) is greatest at the most unfavorable declination is that: (1) the difference is created by weather effects, (2) the weather effects are greatest at low elevation angles, and (3) the unfavorable declination is where the peak elevation angle for a DSN station pass is least.

### C. Effect of Declination

Examination of the data in Figs. 6 and 7 shows the performance advantage of 32 GHz over X-band is a strong function of declination. For 0.95 link reliability, the variation in R1 shown in Fig. 7 is 2.03 dB for Goldstone, 2.53 dB for Canberra, and 3.25 dB for Madrid. These differences illustrate the importance of being able to use the most favorable DSN station location for a given declination. The advantage of being able to use the most favorable DSN station location for each declination is not unique to 32 GHz. During the MM II/CASSINI (1993 launch) encounter (the orbital phase of the mission), which lasts for nearly four years, the spacecraft declination varies between  $19^\circ$  and  $22^\circ$ . At the first Titan encounter after SOI (Saturn orbital insertion), the declination is  $21^\circ$  and the total data return with an X-band link using the

Goldstone DSN 70-m station is 2.7 times that which can be obtained using the Canberra DSN 70-m station.

### D. Effect of DSN 70-m Station Improvements for 32 GHz

Figure 8 shows curves of R1 as a function of declination for the three DSN station locations and the "Baseline," "Passive Improvements," and "Mechanical Compensation" DSN 70-m station 32-GHz antenna gain models. Results for the "Electronic Compensation" DSN 70-m station antenna gain model were omitted because its performance, as shown in Fig. 1, is very nearly the same as that of the "Mechanical Compensation" DSN 70-m station 32-GHz antenna gain model. The results in Fig. 8 are for a 0.95 link reliability and the Slobin/Average weather model. For  $0^\circ$  declination, the separation of the "Baseline" and "Passive Improvement" curves in Fig. 8 is 1.43 dB, independent of DSN station location. For  $0^\circ$  declination, the differences between the "Mechanical Compensation" and "Passive Improvement" curves are 0.76 dB for Goldstone, 0.81 dB for Canberra, and 0.75 dB for Madrid. Remember that this difference includes only a  $-0.07$  dB pointing error differential, which is based on  $0.001^\circ$  32 GHz and  $0.003^\circ$  X-band DSN 70-m station antenna pointing errors. Achieving a  $0.001^\circ$  DSN antenna pointing error is probably not feasible with the "Baseline" antenna. With the current accuracy of about  $0.005^\circ$ , the DSN 70-m station antenna pointing loss at 32 GHz would be about 4.5 dB. At  $0.005^\circ$  pointing error, the DSN 70-m station antenna pointing loss at X-band would be about 0.3 dB.

### E. Effect of Weather Model

Figure 9 differs from the preceding figures in that it is not a comparison of 32-GHz link total data return with X-band link total data return, but a comparison of 32-GHz total data returns for different weather models. There are three sets of curves in Fig. 9 with three curves, for the three DSN station locations, in each set. The sets compare 32-GHz link performance using the Slobin/Best, Slobin/Worst, and Slobin/810-5 weather models with that achieved using the Slobin/Average weather model. Performance with the Slobin/Best weather model is no more than about 1 dB better than performance with the Slobin/Average weather model. Performance with the Slobin/Worst weather model is no more than about 1.2 dB worse than performance with the Slobin/Average weather model. However, performance with the Slobin/810-5 weather model (year average weather) can be as much as 6.5 dB worse than performance with the Slobin/Average weather model. As noted previously, qualitative observations indicate the Slobin/810-5 model appears slightly conservative for X-band links, and extrapolation to 32 GHz would greatly magnify such errors.

## VIII. Conclusions

This report compares X-band 8.5-GHz and 32-GHz link performance on the basis of the number of bits returned during a DSN station pass using a fixed- (one) rate, two-rate, or variable-rate data-rate strategy. For the fixed- or two-rate strategy, use of the best rate or rates is assumed. For each DSN station location, the 32-GHz performance advantage over X-band is plotted as a function of declination. The advantage of this approach is that declination changes slowly with time during a mission. During the four-year MM II/SOTP (1993 launch) encounter period, the declination remains within the  $19^{\circ}$  to  $23^{\circ}$  range. Previously, the ratio of achievable data rate was used to measure the performance advantage of 32 GHz over X-band. This ratio is a function of DSN station elevation angle, which varies over a major part of its possible range during a single DSN station pass.

Figures 6 through 8 show that the performance advantage of 32 GHz over X-band is very dependent on declination and DSN station location. However, examining Fig. 7 one finds that a 8.2-dB advantage can always be obtained for at least one DSN station. Even if Goldstone is not used, a 7.7-dB advantage can be obtained for at least one DSN station. These improvements assume the use of a fixed (one) data rate for each DSN station pass. If multiple rates can be used during each DSN station pass, increasing the complexity of mission operations, Fig. 6 shows that the performance advantage of 32 GHz over X-band can be increased further by 0.2 to 1.0 dB. Figures 6 and 7 assume the use of a mechanically compensated DSN 70-m antenna for 32 GHz and the Slobin/Average weather model. Comparison of the DSN 70-m station "mechanically compensated" and "electronically compensated" 32-GHz antenna gain models shown in Fig. 1 suggests that the performance advantage of 32 GHz over X-band for the electronically compensated DSN 70-m antenna would be similar to the results discussed above for the mechanically compensated DSN 70-m antenna.

The comparison of the 32-GHz link performance using the Slobin/Best, Slobin/Worst, and Slobin/810-5 weather models with 32-GHz link performance using the Slobin/Average weather model in Fig. 9 shows that the selection of the weather model has a very significant impact on 32-GHz link performance. The Slobin/Average weather model was used for the results shown in Figs. 4 through 8. The Slobin/Best, Slobin/Average, and Slobin/Worst weather models are all based on Goldstone K-band radiometer noise temperature measurements at  $30^{\circ}$  elevation angle with extrapolation to the overseas stations and other elevation angles using weather statistics for comparable sites. The Slobin/810-5 weather model is based on models for X-band attenuation from water vapor, clouds, and rain for sites similar to the DSN station locations. The curves in Fig. 9 show that the Slobin/Best and Slobin/Worst weather model cause 32-GHz link performance to vary at most +1.0, -1.2 dB from that for the Slobin/Average weather model. However, with the Slobin/810-5 weather model, the 32-GHz link performance is as much as 6.5 dB worse than that with the Slobin/Average weather model. The conclusion is that either the Slobin/810-5 weather model at 32 GHz is unduly conservative or the Slobin/Best, Slobin/Average, and Slobin/Worst weather models are very optimistic. Qualitative experience suggests the Slobin/810-5 X-band model is slightly conservative, and the method of extrapolating weather effects from X-band to 32 GHz would tend to greatly magnify any error. Finally, it should be noted that the DSN 70-m station 32-GHz antenna gain models and the 32-GHz weather models used in these calculations are very preliminary engineering estimates. The DSN 70-m station clear-weather noise temperature model at both X-band and 32 GHz should be reviewed to reflect improvements expected to be incorporated by the mid-1990s. Because the effect of weather on receiving system noise temperature is much greater for 32 GHz than X-band, an equal decrease in 32 GHz and X-band clear-weather receiving system noise temperature would reduce the performance advantage of 32 GHz over X-band. Most of all, considerable additional attention should be placed on the construction of better 32-GHz weather models than those used in this report.

## Reference

1. Potter, P. D., "64-Meter Antenna Operation at K<sub>a</sub>-Band," *JPL TDA Progress Report 42-57*, March and April 1980, pp. 65-70.

**Table 1. 32-GHz/Baseline antenna gain versus elevation angle  
model fixed losses**

RF losses	Loss, dB
Waveguide loss	-0.088
Forward spillover	-0.132
Rear spillover	-0.013
Illumination	-0.088
Phase	-0.087
Central blockage	-0.044
M No. 1 modes	-0.096
VSWR	-0.044
Mesh loss	-0.009
RF loss Subtotal	-0.601
Mechanical and other losses	
Quadrupod blockage	-0.364
Reflector panels (0.127 mm (0.005 in.))	-0.126
Panel setting (0.381 mm (0.015 in.))	-1.133
Subreflector (0.254 mm (0.010 in.))	-0.504
Thermal (0.254 mm (0.010 in.))	-0.504
Wind (32.2 km/h (20 mph), 0.279 mm (0.011 in.))	-0.609
Mechanical and other loss subtotal	-3.240
Total fixed loss	-3.841

**Table 2. DSN 70-m 32-GHz/Baseline antenna gain reduction from gravity-induced surface distortions**

Elevation angle, deg	$L_{GV}$ , dB
6.	-2.495
10.	-1.907
20.	-0.889
30.	-0.316
40.	-0.086
50.	-0.215
60.	-0.514
70.	-1.219
80.	-1.979
90.	-2.940

**Table 3. DSN 70-m station 32-GHz system temperature increase from weather at 30° elevation angle**

Cumulative probability	Noise temperature increase, K					
	Goldstone			Canberra and Madrid		
	Best	Average	Worst	Best	Average	Worst
0.0	10.0	10.0	10.0	10.5	10.5	10.5
0.2	13.5	14.0	14.5	16.5	17.5	18.5
0.5	17.0	18.0	19.0	23.0	25.0	27.0
0.7	19.0	20.5	22.0	26.0	28.5	32.0
0.8	21.0	23.0	25.0	29.0	32.0	37.0
0.9	24.0	26.0	30.0	33.0	38.0	45.0
0.95	27.0	31.0	35.0	37.0	46.0	55.0
0.98	32.0	37.0	43.0	46.0	62.0	120.0
0.99	36.0	44.0	53.0	57.0	120.0	180.0
0.995	41.0	53.0	69.0	75.0	150.0	220.0
0.998	51.0	80.0	175.0	180.0	215.0	260.0
0.999	65.0	120.0	240.0	260.0	270.0	279.0

**Table 4.** Sample X-band link performance estimate

COMPUTER PROGRAM NO. 530-0P                    09-26-86

TELECOMMUNICATION SYSTEM DOWNLINK PERFORMANCE ESTIMATE

TEST CASE FOR K-BAND/X-BAND COMPARISON (1-9-86)

X=BAND/10.7 W XSSPA  
 X=BAND/3.67 H BODY-FIXED HGA, 0-18 DEG. POINTING ERROR  
 DBN 70 H STATION/X-BAND/DIMPLEXED/PE = 0.003 DEG.  
 CANBERRA/EXTRAPOLATION FROM K-BAND NOISE TEMPERATURE MEASUREMENTS/AVERAGE  
 DSN BLOCK III RECEIVER/10.8 HZ BANDWIDTH MODE  
 REED-SOLOMON/VITERBI CHANNEL, PB = 1.E-6

ACHIEVABLE DATA RATE AT 10 AU						
	DESIGN VALUE	TOLERANCE	DIST	MEAN	VARIANCE	
TRANSMITTING SYSTEM PARAMETERS						
1. RF POWER OUTPUT	40.29 DBH	.1.00 DB	-1.00 DB	1	40.29 DBM	.1667 DB**2
2. TRANSMITTER CIRCUIT LOSS	-.45 DB	.20 DB	.20 DB	U	-.45 DB	.0133 DB**2
3. ANTENNA CIRCUIT LOSS	-.30 DB	.10 DB	.10 DB	U	-.30 DB	.0033 DB**2
4. ANTENNA GAIN (REFERENCED TO CP)	48.80 DB	.50 DB	.50 DB	T	48.80 DB	.0417 DB**2
5. ANTENNA ELLIPTICITY	1.70 DB	.00 DB	.00 DB	T	1.00 DB	.0000 DB**2
6. ANTENNA POINTING LOSS	-1.00 DB	.00 DB	.00 DB	T	1.00 DB	.0000 DB**2
PATH PARAMETERS						
6. SPACE LOSS						
RANGE = 10.0000 AU	294.44 DB					
FREQUENCY = 8.4150 GHZ						
RECEIVING SYSTEM PARAMETERS						
7. POLARIZATION LOSS						
8. ANTENNA GAIN (FOR MATCHED POLARIZATION)	73.76 DB	.05 DB	-.07 DB	U	73.76 DB	.0011 DB**2
9. ANTENNA POINTING LOSS	1.00 DB	.20 DB	-.50 DB	U	1.00 DB	.1200 DB**2
10. NOISE SPECTRAL DENSITY	-.163.83 DBM/HZ	.00 DB	.00 DB	U	-.163.85 DBM/HZ	.0000 DB**2
SYSTEM NOISE TEMPERATURE	29.98 K	.41 DB	-.46 DB	T	30.00 K	.0317 DB**2
ZENITH NOISE TEMPERATURE	25.00 K	3.00 K	3.00 K			
ADD. FOR ELEVATION ANGLE	4.98 K	.00 K	.00 K			
ELEVATION ANGLE	30.0000 DEG					
11. AVAILABLE PT/NO						
TELEMETRY PERFORMANCE ESTIMATE						
12. RANGING CARRIER SUPPRESSION	0.00 DB	.00 DB		T	0.00 DB	.0000 DB**2
13. REQUIRED PT/NO (RANGING OFF)	49.00 DB.HZ	.50 DB	-.50 DB	U	49.00 DB.HZ	.0033 DB**2
DATA RATE	3C699.35442 BPS					
REQUIRED EB/NO	a.0000 DB					
REQUIRED CARRIER MARGIN	18.3500 DB					
MODULATION LEVEL (RMS)	60.0000 DEG					
14. PERFORMANCE MARGIN						
15. LINK RELIABILITY SIGMA						

ORIGINAL PAGE IS  
OF POOR QUALITY

Table 5. Sample K-band link performance estimate

COMPUTER PROGRAM NO. 5300P	09-26-86
TELECOMMUNICATION SYSTEM DOWNLINK PERFORMANCE ESTIMATE	
TEST CASE FOR K-BAND/X-BAND COMPARISON (1-9-86)	
K-BAND/10.7 W KASPA	
K-BAND/3.67 M BODY-FIXED HGA, 0.047 DEG. POINTING ERROR	
DSN 70 M STATION/K-BAND/MECHANICAL COMPENSATION/PE = 0.001 DEG.	
CANBERRA/EXTRAPOLATION FROM K-BAND NOISE TEMPERATURE MEASUREMENTS/AVERAGE	
DSN BLOCK III RECEIVER/10.8 Hz BANDWIDTH MODE	
REED-SOLOMON/VITERBI CHANNEL, PB = 1.E-6	
ACHIEVABLE DATA RATE AT 10 AU	
TRANSMITTING SYSTEM PARAMETERS	
1. RF POWER OUTPUT	40.29 DBM
2. TRANSMITTER CIRCUIT LOSS	.45 DB
3. ANTENNA CIRCUIT LOSS	.30 DB
4. ANTENNA GAIN (REFERRED TO CP)	60.40 DB
ANTENNA ELLIPTICITY	1.70 DB
5. ANTENNA POINTING LOSS	-1.00 DB
PATH PARAMETERS	
6. SPACE LOSS RANGE = 10.0000 AU FREQUENCY = 32.0000 GHZ	-306.04 DB
RECEIVING SYSTEM PARAMETERS	
7. POLARIZATION LOSS	.01 DB
8. ANTENNA GAIN (FOR HATCHED POLARIZATION)	84.89 DB
ANTENNA ELLIPTICITY	1.00 DB
9. ANTENNA POINTING LOSS	.17 DB
10. NOISE SPECTRAL DENSITY SYSTEM NOISE TEMPERATURE	163.66 DBM/HZ 31.20 K
ZENITH NOISE TEMPERATURE	23.15 K
ADD. FOR ELEVATION ANGLE ELEVATION ANGLE	.05 K 30.0000 DEG
11. AVAILABLE PT/NO	
TELEMETRY PERFORMANCE ESTIMATE	
12. RANGING CARRIER SUPPRESSION	.00 DB
13. REQUIRED PT/NO (RANGING OFF)	56.99 DB.HZ
DATA RATE	193253.2285 BPS
REQUIRED EB/NO	4.0000 DB
REQUIRED CARRIER MARGIN	18.3500 DB
MODULATION LEVEL (RMS)	80.0000 DEG
14. PERFORMANCE MARGIN	.50 DB
15. LINK RELIABILITY SIGMA	.9500 DB .68 DB

ORIGINAL PAGE IS  
OF POOR QUALITY

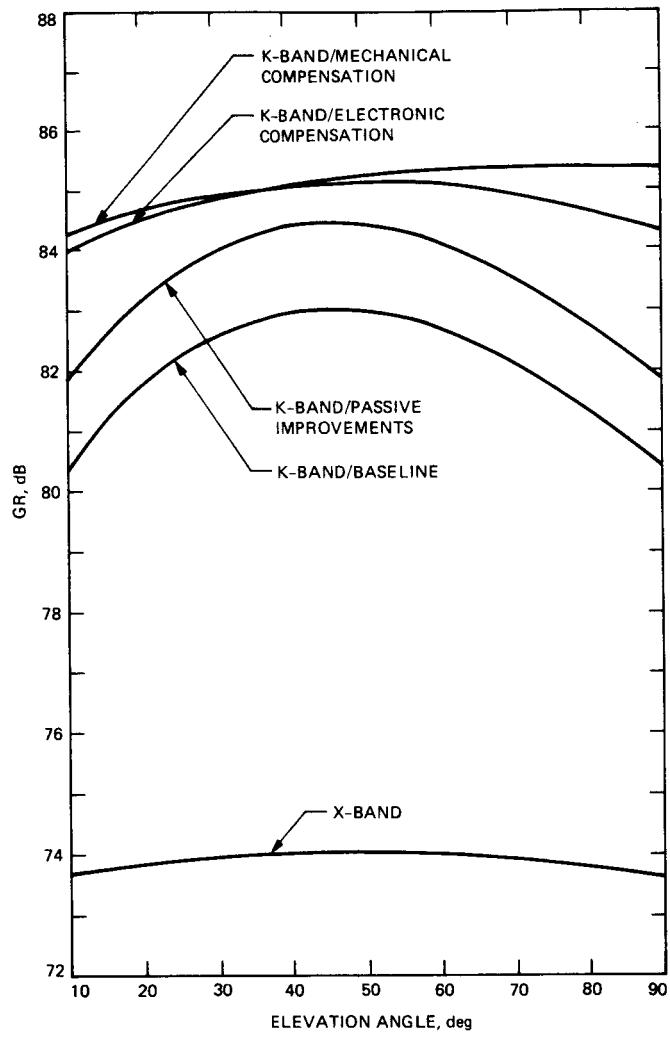


Fig. 1. DSN 70-m station X-band and K-band antenna gain models

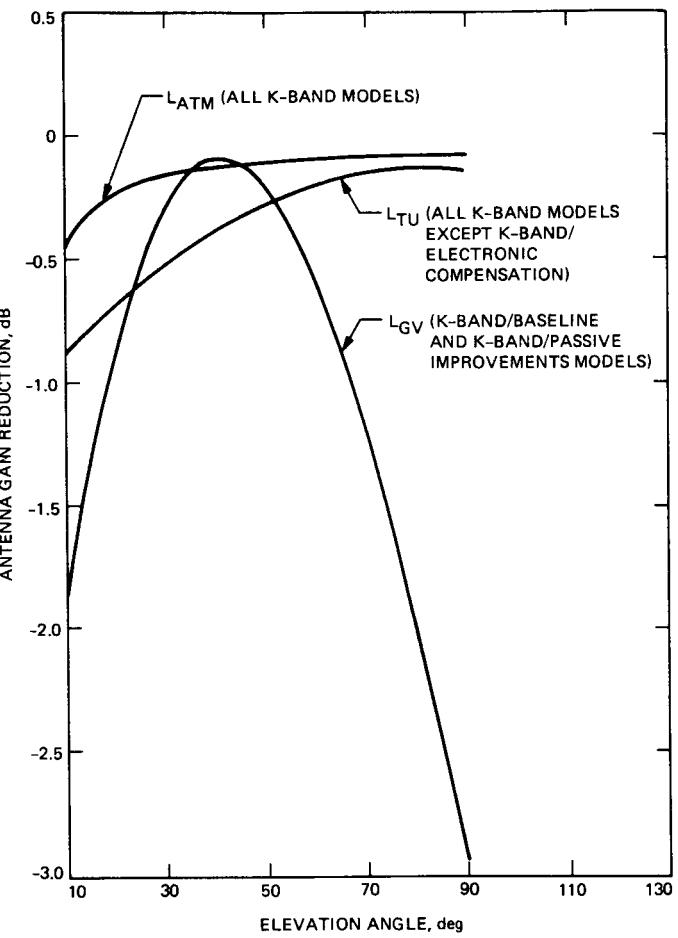


Fig. 2. DSN 70-m station K-band antenna gain reduction from gravity-induced surface distortions, atmospheric turbulence, and atmospheric attenuation

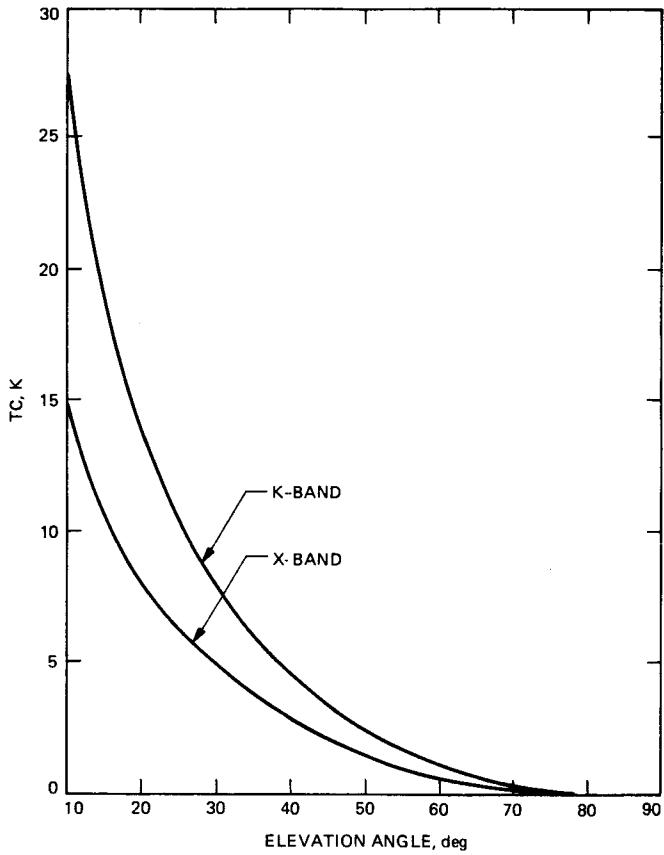


Fig. 3. DSN 70-m station clear-weather receiving system noise temperature increase for nonzenith elevation angles

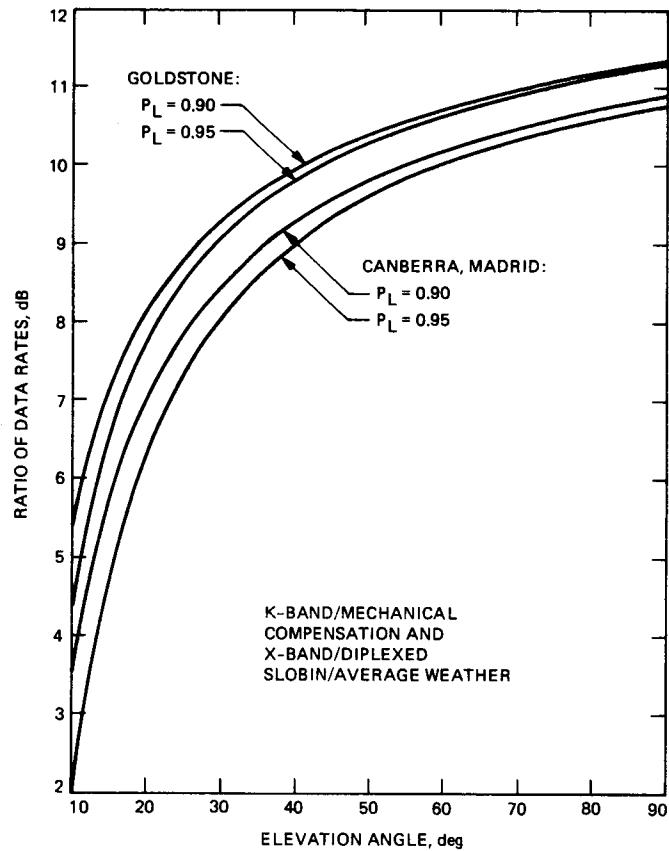
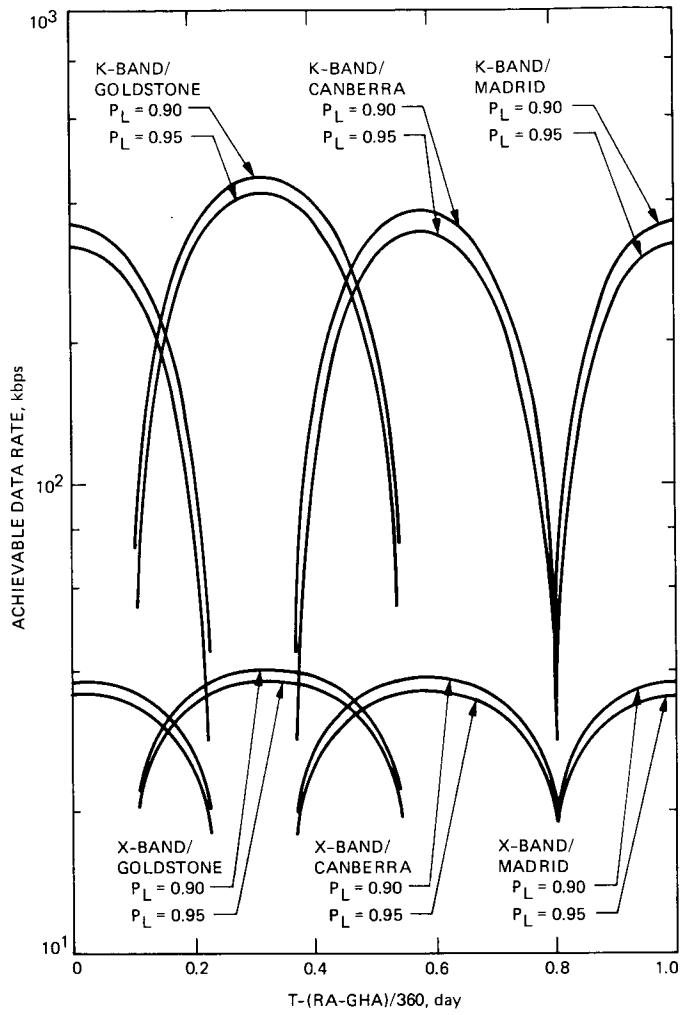
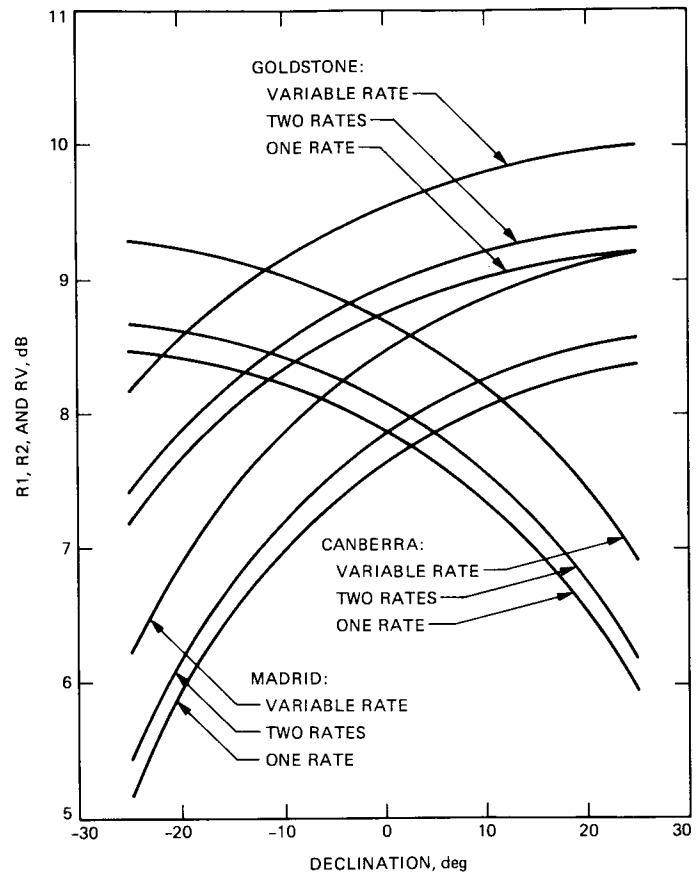


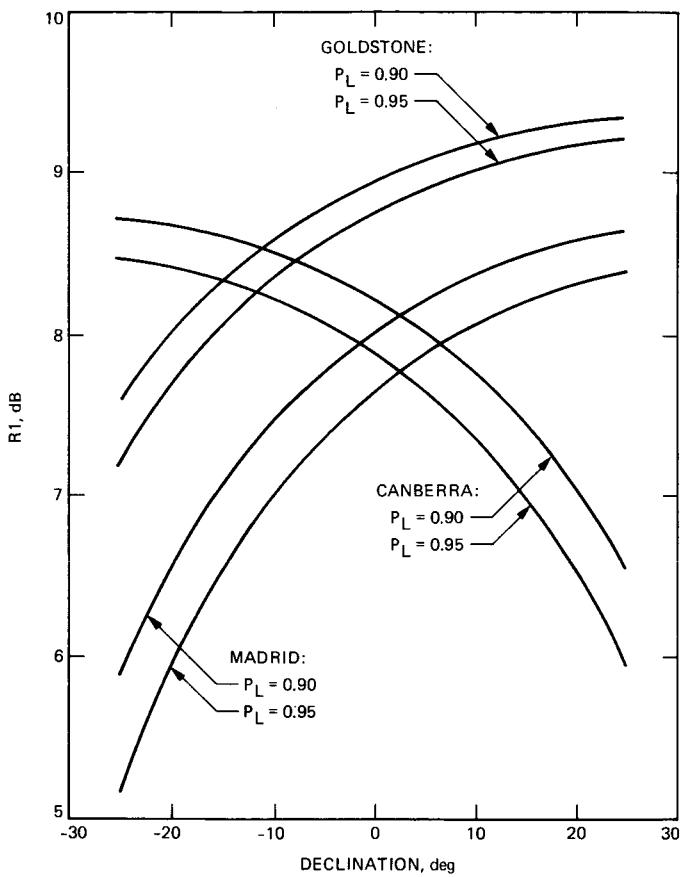
Fig. 4. Ratio of K-band to X-band achievable data rate as a function of elevation angle with DSN station location and link reliability as the curve parameters



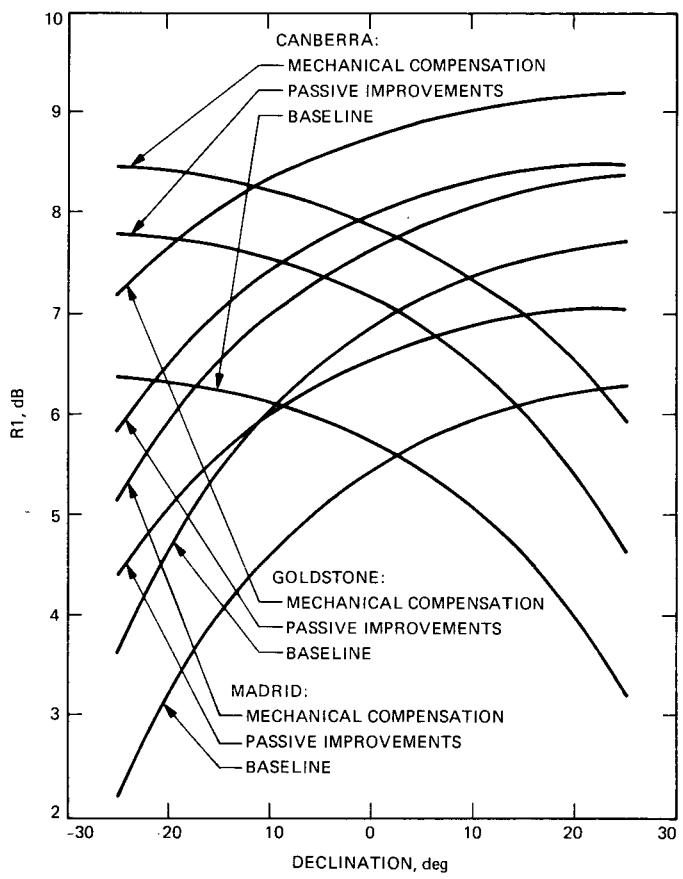
**Fig. 5. Comparison of X-band and K-band achievable data rate as a function of time for a one-day period at 0° declination for a DSN 70-m station K-band antenna with mechanical compensation and Slobin/Average weather**



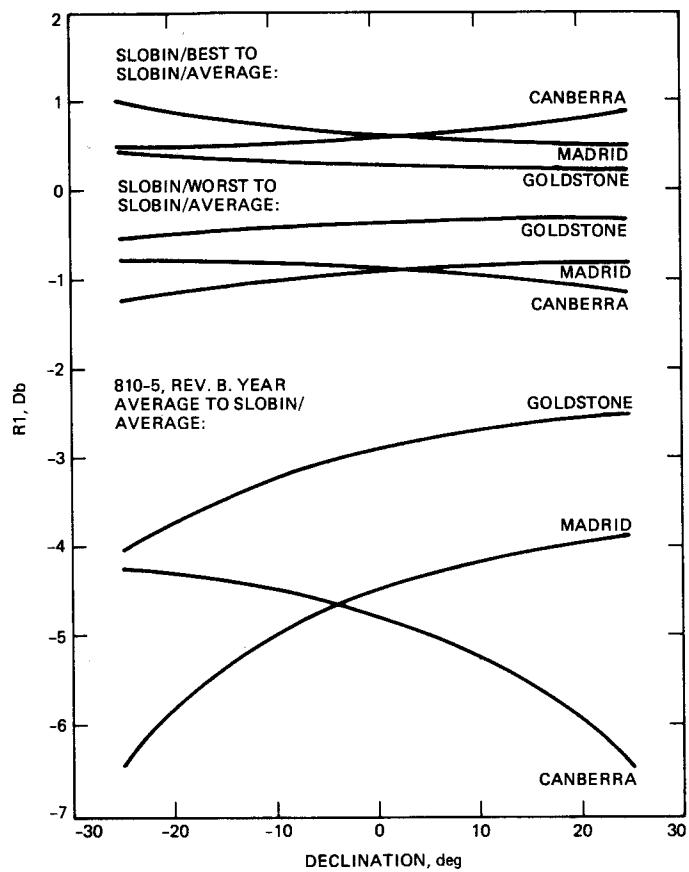
**Fig. 6. Ratio of K-band to X-band total data return per pass as a function of declination for a fixed-rate, two-rate, and variable-rate data-rate strategy and for a DSN 70-m station K-band antenna with mechanical compensation, Slobin/Average weather, and 0.95 link reliability**



**Fig. 7.** Ratio of K-band to X-band total data return per pass as a function of declination for a fixed data-rate strategy and for a DSN 70-m station K-band antenna with mechanical compensation and Slobin/Average weather



**Fig. 8.** Ratio of K-band to X-band total data return per pass as a function of declination for a fixed data-rate strategy, Slobin/Average weather, and 0.95 link reliability with DSN station location and the DSN 70-m station K-band antenna gain model as the curve parameters



**Fig. 9. Comparison of K-band total data return per pass using the Slobin/Best, Slobin/Worst, and Slobin/810-5 weather models with that using the Slobin/Average weather model for a DSN 70-m station K-band antenna with mechanical compensation and 0.95 link reliability**

*Dg - 32*  
**N87-15338**

*P-11*

# **Long-Term Amplitude and Phase Stability of the 400-kW 2.115-GHz Transmitter**

**D. J. Hoppe and A. M. Bhanji**  
Radio Frequency and Microwave Subsystems Section

*Results of recent measurements of the long-term phase, amplitude and group delay stability of the 400-kW S-band (2.115-GHz) transmitter are reported. Various control parameters which are responsible for many of the observed instabilities are identified. Further tests to identify the parameters responsible for the remaining instabilities are suggested.*

## **I. Introduction**

The results of measurements which were taken recently to determine the stability of the 400-kW S-band (2.115-GHz) transmitter over a typical track time are described. Of particular importance are the amplitude, phase, and group delay stability of the transmitter system. In addition to showing that the transmitter meets the specifications which have been placed on it, the tests also show which of the various control parameters of the transmitter (e.g., beam voltage, temperature) are responsible for the observed instabilities. A brief description of the experiment is given, and plots of some of the more important results are presented. The results are interpreted, and some suggestions for improving the transmitter's stability are discussed. Further tests to verify these improvements are described.

was controlled by an HP Series 200 computer, and used to drive the 400-kW transmitter. Measurements of the transfer function through the transmitter and several other transmitter parameters were monitored through an HP data acquisition unit and the existing monitoring system. Data was acquired at regular intervals and stored on a disk for later processing.

The transmitter was operated continuously for approximately a 4-h period with no operator intervention except in the case of an automatic emergency shutdown (e.g., crowbar). Data was gathered at approximately 30-s intervals throughout the 4-h period. The gain and phase through the transmitter system were measured by the network analyzer at 2.114, 2.115, and 2.116 GHz for each data collection time. The phase and amplitude data taken by the network analyzer was obtained by averaging 512 measurements at each frequency of interest. In addition, the existing transmitter monitoring system was interfaced to the HP computer through an HP data collection system. This allowed for the monitoring of a number of auxiliary transmitter parameters throughout the experiment. In particular, klystron beam voltage, beam current, filament current, filament voltage, vacuum current, magnet

## **II. Experimental Setup**

The interconnection of the various test equipment used to monitor the transmitter parameters during the simulated track period is depicted in Fig. 1. An HP 8510 network analyzer

current, body current, and collector current were monitored throughout the simulated track. The inlet temperature of the coolant used to cool the klystron and water load was also measured using a separate thermometer. Transmitter output power at 2.115 GHz, reflected power at 2.115 GHz, and drive power to the klystron at 2.115 GHz were also recorded. Using the phase data collected at 2.114 and 2.116 GHz, the group delay at 2.115 GHz was calculated. The transmitter was operated in a saturated mode with a nominal output power of 400 kW.

### III. Results

Figures 2 through 5, 6(a), and 7(a) summarize the long-term behavior of some of the transmitter parameters. Figure 6(a) shows transmitter output power as a function of time over the 250-min experiment time. The sudden drop in output power at approximately sample number 150 was caused by an automatic firing of the crowbar system. The transmitter was then brought up to full power as soon as possible, and the experiment was resumed. The starting value for the output power was 397.2 kW (saturated). The maximum power observed during the experiment was about 15 kW above the starting value; the minimum was about 6 kW below the initial output power. These values are well within the specified limits for the transmitter, which are +135 kW and -106 kW ( $\pm 1.3$  dB).

The phase stability of the transmitter system, from the output port of the HP 8510 network analyzer to the input port, is depicted in Fig. 7(a). Maximum phase changes of +12 deg and -6 deg were observed, if the extraneous data taken during the time the transmitter was off is neglected. Once again, these values are well within the allowable  $\pm 550$  deg specified for the transmitter.

Figure 2 shows the change in group delay through the transmitter system as a function of time. As was explained earlier, the delay was determined by using phase data at 2.114 GHz and 2.116 GHz to approximate the slope of the phase curve at 2.115 GHz. Some smoothing has been applied to the data before plotting. Peak deviations of +0.8 ns and -1.5 ns observed, once again well within the  $\pm 3.3$ -ns specification for the system.

The final plots show the three control parameters which are believed to be most important in determining the phase and amplitude stability of the transmitter. Klystron beam voltage is shown in Fig. 3. No downward drift from the initial value of 57.6 kV was observed, but an increase of about 750 V can be seen. This corresponds to a regulation of about 1.3 percent. Subsequently, an inspection of the beam voltage regulation circuitry revealed an incorrect resistor value, which caused the reference voltage to behave abnormally. As will be

seen later, much of the observed instability in gain and phase can be attributed to this parameter.

The inlet coolant temperature was also monitored and is depicted in Fig. 4. The sharp decrease in temperature near sample 150 is associated with the klystron trip-off, and an exponential rise in temperature after the transmitter was restarted can be clearly seen.

The final control parameter that is of interest is the drive power to the klystron, which is provided by a 10-W, Class C amplifier. The initial drive power of 5.91 W had a peak increase of about 80 mW, and a decrease of about 95 mW, as can be seen in Fig. 5. It will be shown in the next section that since the klystron was operated in a saturated mode, these small drive level changes do not significantly affect the amplitude stability of the transmitter, and the phase changes induced through AM-to-PM conversion in the saturated klystron appear to be masked by other effects.

### IV. Discussion

In addition to verifying that the transmitter system meets its stability requirements, it is also of interest to determine which control parameters are responsible for the observed instabilities, and to what extent. The effects of the various control parameters (i.e., beam voltage, temperature, drive power, etc.) may be summarized by two simple equations:

$$P_{out}(t) = P_{out}(t_0) + K_1 [V_B(t) - V_B(t_0)] \\ + K_2 [T(t) - T(t_0)] \\ + K_3 [P_{drive}(t) - P_{drive}(t_0)] + \dots \quad (1)$$

$$\theta(t) = \theta(t_0) + K_4 [V_B(t) - V_B(t_0)] \\ + K_5 [T(t) - T(t_0)] \\ + K_6 [P_{drive}(t) - P_{drive}(t_0)] + \dots \quad (2)$$

where

$t$  = time

$t_0$  = starting time of the experiment

$P_{out}(t) - P_{out}(t_0)$  = change in output power from initial value

$\theta(t) - \theta(t_0)$  = change in phase from initial value

$V_B(t) - V_B(t_0)$  = change in beam voltage from initial value

$T(t) - T(t_0)$  = change in coolant temperature from initial value

$P_{\text{drive}}(t) - P_{\text{drive}}(t_0)$  = change in klystron drive power from initial value

The factors  $K_1 - K_3$  are the pushing factors for power out, and  $K_4 - K_6$  are the pushing factors for output phase. These are the numbers we wish to determine.

The results for output power are shown in Fig. 6. Figure 6(a) shows the raw output power data. It is well known that the beam voltage is an important parameter in determining klystron output power. A preliminary value for  $K_1$  in kW/kV was determined from published tube data, and further refined by trial and error. The final value obtained was about 20 kW/kV, and the results for the change in power out with this beam voltage effect removed are shown in Fig. 6(b).

An exponential effect is clearly visible after the transmitter was restarted. This effect is related to the coolant temperature depicted in Fig. 4. The temperature effects were removed by trial-and-error fitting of the curves, and a value of -0.9 kW/ $^{\circ}\text{C}$  was determined for  $K_2$ . The power out data with both the beam voltage and coolant effects removed is shown in Fig. 6(c).

Upon comparing Fig. 6(c) and Fig. 5, it does not appear that the remaining output power instability is related to drive power. Filament voltage and current are also known to have an effect on output power, but, like drive power, no correlation between the filament data and the remaining output power instabilities could be seen. The remaining 4-kW (1 percent) drop in power may be related to changes in the monitoring system, ambient temperature, or some other control parameter that was not monitored.

A similar procedure was used to determine the phase pushing factors  $K_4 - K_6$ . The measured phase data is shown in Fig. 7(a). A beam voltage effect of about 20 deg/kV was then removed from this data and is plotted in Fig. 7(b). Once again, a temperature effect is clearly visible after the transmitter is turned back on. Figure 7(c) shows the output phase with these effects removed. The value determined for  $K_5$  was -0.8 deg phase/ $^{\circ}\text{C}$ . Once again, the remaining phase instability

does not seem to be correlated with klystron drive power or filament voltage and current, and a value for  $K_6$  could not be determined.

Two major effects remain in the phase data after beam voltage and coolant temperature effects are removed: a long-term drift of about 6 deg, and two phase steps, one early in the data (samples 10-15) and one at about samples 225-325. These phase steps appear to be correlated with the steps in group delay appearing in Fig. 2. The long-term drift may be caused by changes in ambient temperature, or instrument drift. The steps in phase and group delay may be attributed to changes in drive to the 10-W, Class C amplifier or in its power supply, but do not appear to be related to temperature or any klystron parameter. Further tests of the drive system are required to determine the cause of the observed steps in phase and group delay.

Using the phase data that was collected, calculations were also performed to approximate the fractional frequency stability for the *open-loop* transmitter system. A summary of the equations used is given in the appendix. In particular, a value of  $4.9 \times 10^{-15}$  for  $\Delta f/f$  was determined for an averaging time of 990 s. Figure 8 shows the calculated values of  $\Delta f/f$  for various averaging times ( $\tau$ ) for the open-loop transmitter.

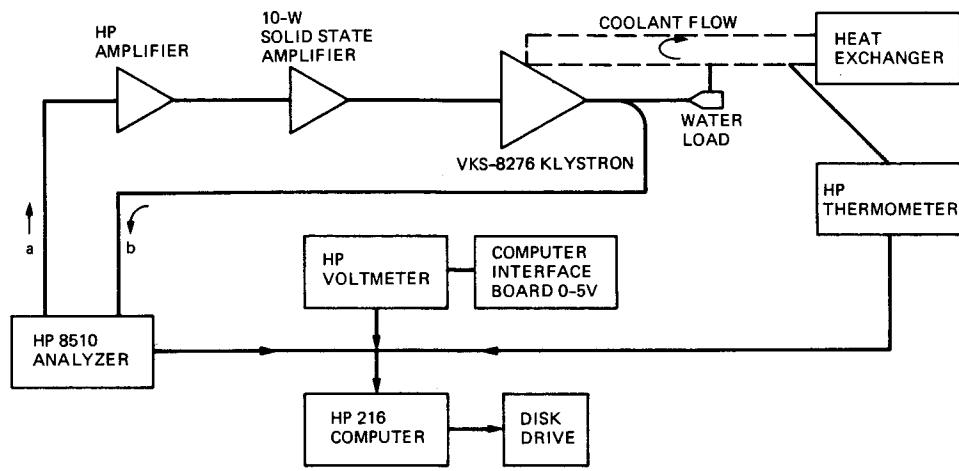
Other control parameters such as filament voltage and current and magnet current were also monitored, but do not appear to be nearly as important as beam voltage and coolant temperature in determining transmitter stability.

## V. Conclusions

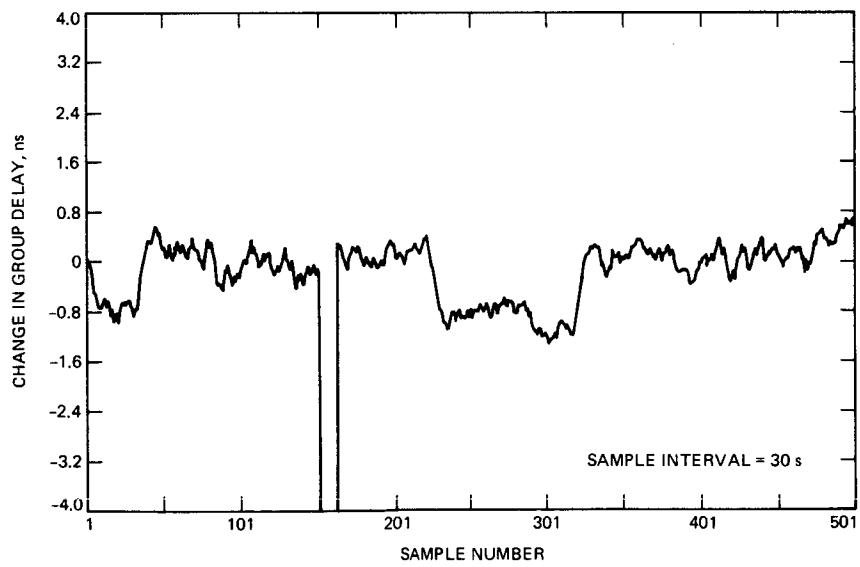
In conclusion, the long-term stability tests have shown that the 400-kW transmitter easily meets its specifications for amplitude, phase, and group stability. Two major contributors to the observed instabilities have been determined: (1) the beam voltage with pushing factors of 20 kW/kV and 20 deg/kV at saturation, and (2) the inlet coolant temperature at -0.9 kW/ $^{\circ}\text{C}$  and -0.8 deg/ $^{\circ}\text{C}$ . Further work is required to determine the contributors to the remaining amplitude, phase, and group delay variations. In particular, a more detailed look at the system of amplifiers driving the klystron would probably be beneficial. If future missions or experiments require greater transmitter stability than was measured during this test, improvements in beam voltage and coolant temperature regulation should be considered.

## **Acknowledgments**

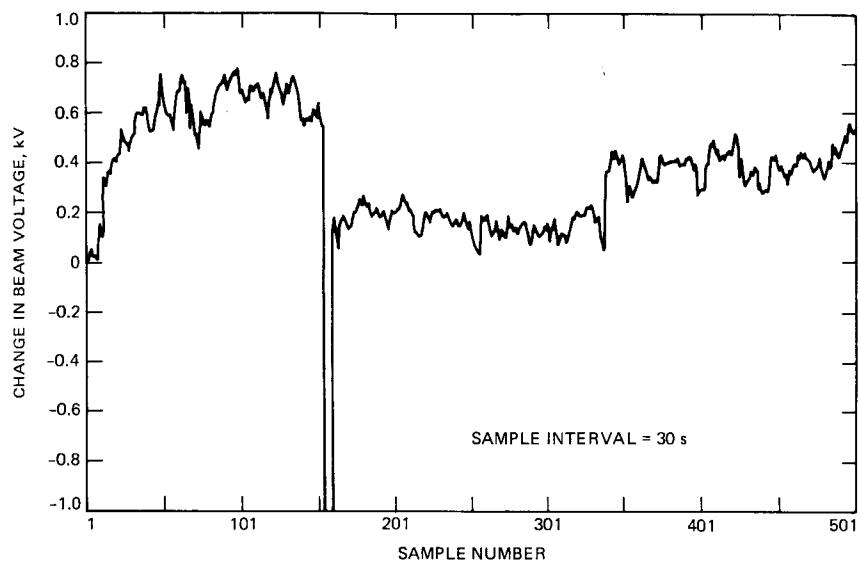
The authors acknowledge the assistance of Marlyn Gregg, John Daeges, and Phil Boss in performing the measurements and interpreting the data, and Roger Meyer of the Communications Systems Research Section for discussions regarding the fractional frequency results.



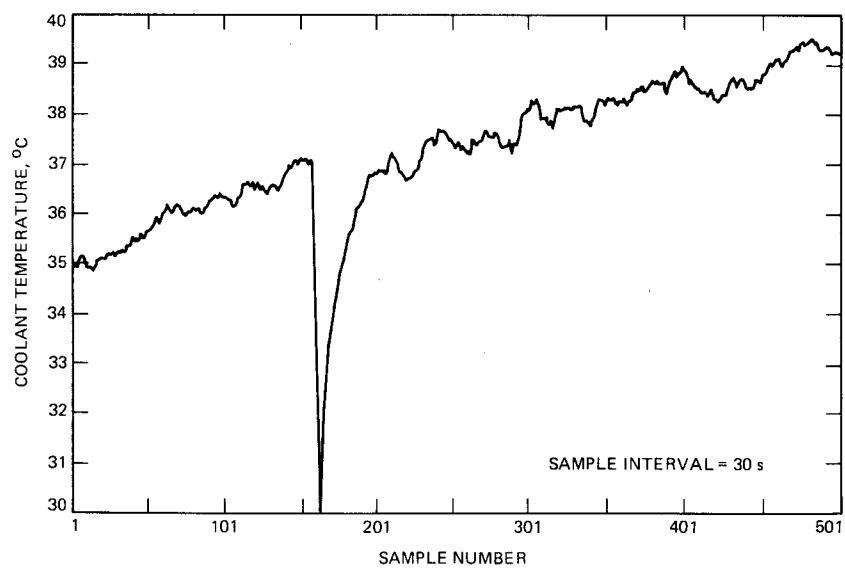
**Fig. 1. Test setup**



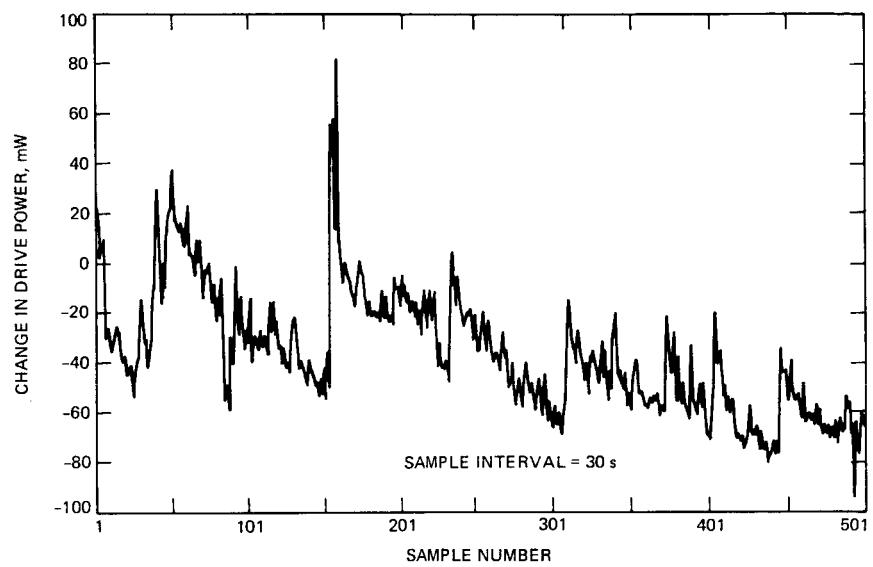
**Fig. 2. Change in group delay versus time**



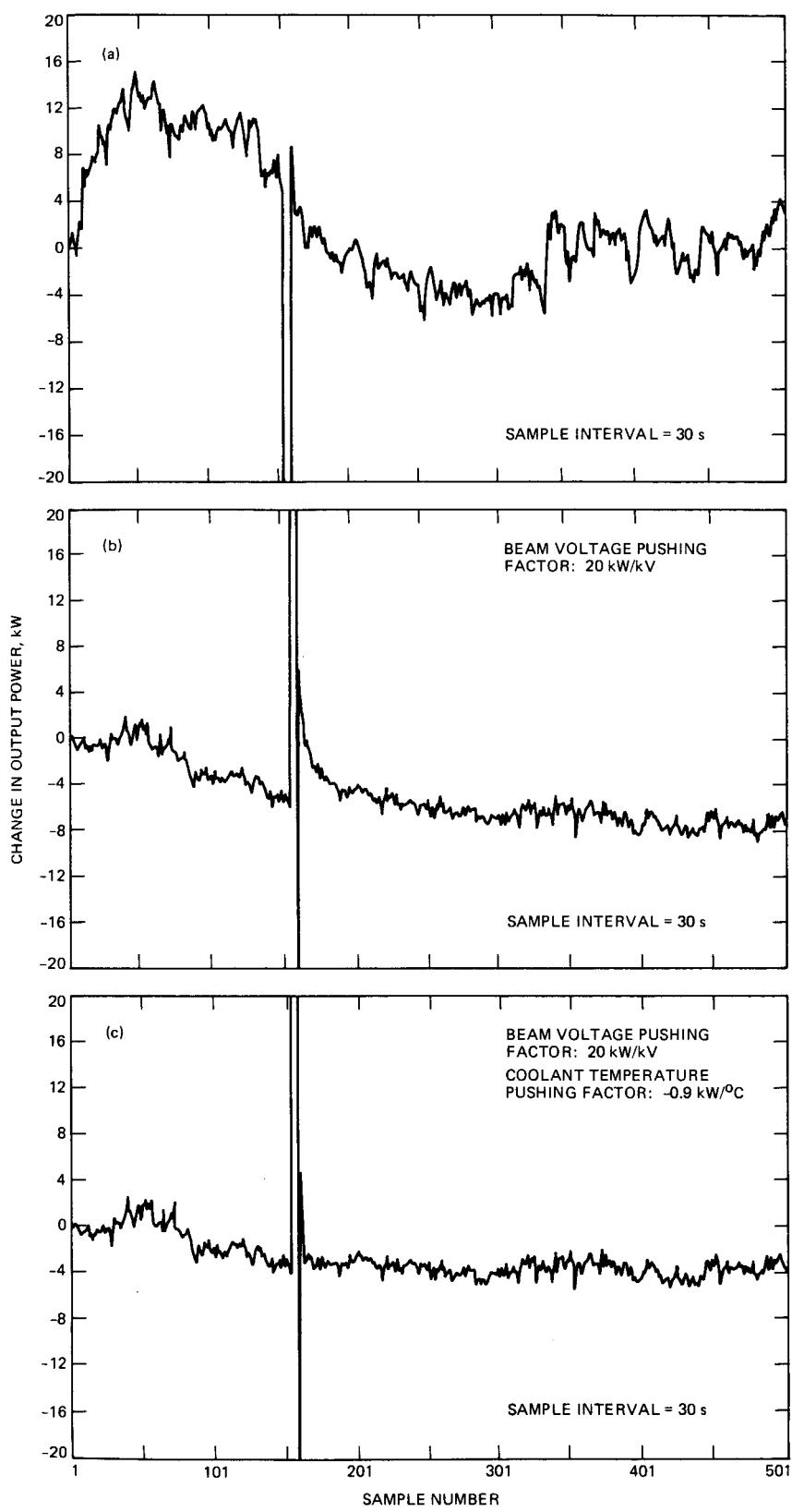
**Fig. 3. Change in beam voltage versus time**



**Fig. 4. Coolant temperature versus time**



**Fig. 5. Change in drive power versus time**



**Fig. 6. Change in output power: (a) uncorrected, (b) beam voltage corrected,  
(c) beam voltage and temperature corrected**

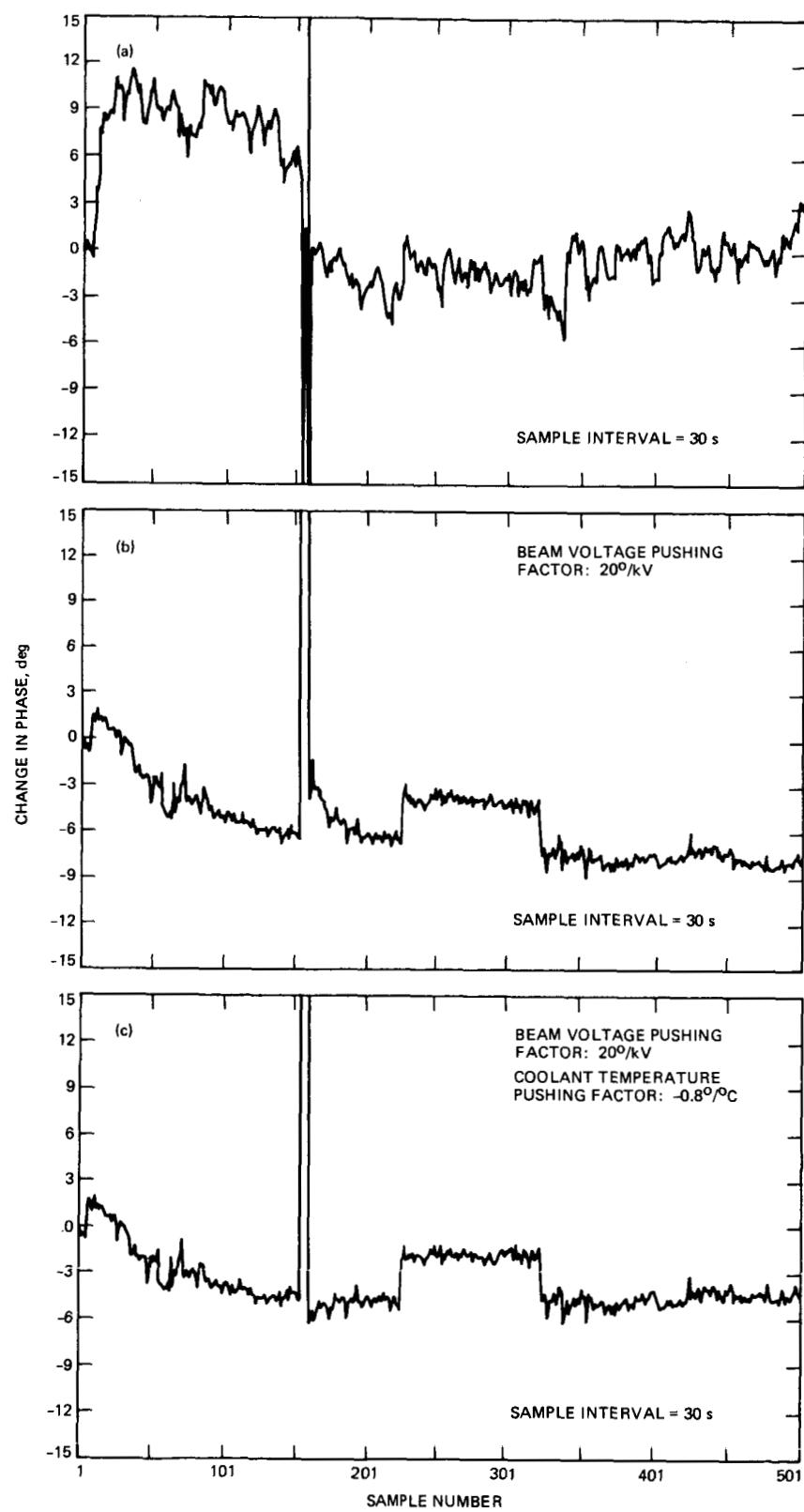
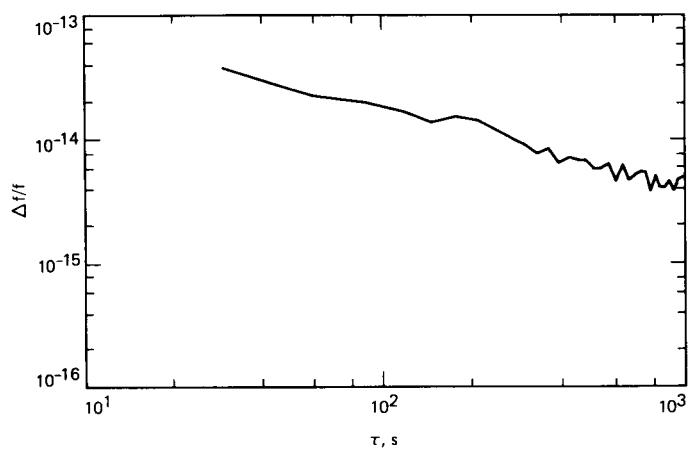


Fig. 7. Change in phase: (a) uncorrected, (b) beam voltage corrected, (c) beam voltage and temperature corrected



**Fig. 8. Fractional frequency stability**

## Appendix

### Calculation of Fractional Frequency Stability From Measured Phase Data

For a general signal with some phase instability, we have

$$V = A \cos [\omega_0 t + \phi(t)] \quad (\text{A-1})$$

The instantaneous phase in radians is

$$\theta(t) = \omega_0 t + \phi(t) \quad (\text{A-2})$$

and the instantaneous frequency is

$$f(t) = \frac{1}{2\pi} \frac{d\theta(t)}{dt} = \frac{1}{2\pi} [\omega_0 + \phi'(t)] \quad (\text{A-3})$$

Define

$$\Delta f(t) = f(t) - f_0 \quad (\text{A-4})$$

$$f(t) = \frac{\phi'(t)}{2\pi} \quad (\text{A-5})$$

The instantaneous fractional frequency fluctuation is

$$\sigma(t) = \frac{\Delta f(t)}{f_0} = \frac{\phi'(t)}{2\pi f_0} \quad (\text{A-6})$$

Next, we find the average value of  $\sigma_\tau(t)$  over an interval of seconds:

$$\bar{\sigma}_\tau(t) = \frac{1}{2\pi f_0} \frac{1}{(t_1 + \tau - t_1)} \int_{t_1}^{t_1 + \tau} \phi'(t) dt \quad (\text{A-7})$$

$$\bar{\sigma}_\tau(t) = \frac{1}{2\pi f_0} \tau [\phi(t_1 + \tau) - \phi(t_1)] \quad (\text{A-8})$$

Equation (A-8) may be evaluated for discrete values of  $\tau$ , namely, integer multiples of the sample time (30 s).

Equation (A-8) is evaluated for each interval of seconds in the data set, and the Allan variance of these evaluations of Eq. (A-8) is found. This process is repeated for each value of  $\tau$ , and the results are usually plotted on a log-log scale as  $\log(\sigma)$  vs  $\log(\tau)$ .

D10-35

N87-15339

TDA Progress Report 42-87

July-September 1986

211501  
P/M

## DSN Microwave Antenna Holography

### Part II: Data Processing and Display of High-Resolution Effective Maps

D. J. Rochblatt

Radio Frequency and Microwave Subsystems Section

Y. Rahmat-Samii and J. H. Mumford

Spacecraft Telecommunications Equipment Section

*This article presents the results of a recently completed computer graphic package for the process and display of holographically recorded data into effective aperture maps. The term "effective maps" (labelled "provisional" on the holograms) signifies that the maps include contributions of surface mechanical errors as well as other electromagnetic factors (phase error due to feed/subreflector misalignment, linear phase error contribution due to pointing errors, subreflector flange diffraction effects, and strut diffraction shadows). While these maps do not show the true mechanical surface errors, they nevertheless show the equivalent errors, which are effective in determining overall antenna performance. Final steps to remove electromagnetic pointing and misalignment factors are now in progress. The processing and display of high-resolution effective maps of a 64m antenna (DSS 63) are presented.*

#### I. Introduction

The fundamentals of the microwave holography technique have been presented by Rahmat-Samii in Refs. 1 and 2, to which the reader is referred for details on mathematical formulations, numerical simulations, application of an iterative algorithm, etc. Additionally, the basic principles of the measurement aspects have been outlined in an earlier TDA Progress Report (Ref. 3), which is considered as Part I of this article. Microwave holography is expected to be a key technique for achieving outstanding DSN 8.4-GHz performance and enabling future DSN capability at even shorter wavelength, such as the 32/34-GHz bands recently allocated for deep space telemetry.

The NASA Deep Space Network is currently upgrading its three 64m antennas to 70m to achieve a microwave gain increase of 1.9 dB at 8.4 GHz (X-band). This performance improvement is applicable to all future deep space missions and to solar system radar, as well as to more immediate support for the 1989 Voyager 2 encounter with Neptune. The precise alignment of the reflector panels and precise location of the subreflector are the final key factors in achieving the 1.9-dB gain increase at X-band, and are essential for future acceptable performance at 32 GHz.

More than 3800 m<sup>2</sup> of area is associated with each 70m antenna, comprising 1272 individual panels and 5472 adjusters.

These panels must be aligned to a precision of 0.25-mm (0.01-in.) rms. In the past, antenna panel alignment was accomplished using optical (theodolite) techniques. These techniques will still be used to provide an initial surface accuracy, but to achieve the final precise panel alignment and precise subreflector position, the microwave holography technique will be applied to advantage (Refs. 1 and 2).

Holography, "total recording," acquires phase and amplitude raster-scan patterns of the antenna far-field angular response (Refs. 1 through 3). The holography metrology is based on interferometrically connecting a reference antenna to the large test antenna and digitally recording the test antenna amplitude and phase response. This is done by continuously scanning the test antenna against a geosynchronous satellite, following a two-dimensional grid. The angular extent (or number of sidelobes) of the response that must be acquired is inversely proportional to the size of the desired resolution cell in the processed holographic maps (Ref. 1). At 11.45-GHz measurement frequency, resolution cells of 0.4m diameter have been achieved, providing several resolution cells per individual reflecting panel. An Inverse Fast Fourier Transform (IFFT) algorithm is then used to obtain the desired information, consisting of the test antenna aperture amplitude and phase response (Ref. 1). From the aperture phase response, the "surface error map" is calculated; the amplitude response ("surface current map") is directly displayed. It is the information in the "surface error map" that is used to calculate the adjustments of the individual panels in an overall main reflector best-fit reference frame. The "surface current map" provides valuable information about the energy pattern distribution in the antenna aperture. JPL engineering has developed the data processing and display capability to produce "effective maps" on a portable HP-1000 computer. The short-term goal is to develop a capability to display a "final map" and panel adjustment correction listing. Thus far, a contractor (Eikontech, Ltd.) has supplied the digitized raw data of the test antenna far-field amplitude and phase response. The long-term goal is to develop (or acquire) the data recording as well as processing and display capability.

## II. High-Resolution Test Results

Figures 1 and 2 are black-and-white presentations of the 14-color computer displays of effective maps of the DSS 63 64m antenna. (The color version may be ordered from the JPL Photo Lab. Reprint order numbers for Figures 1 and 2 are JPL 248AC and JPL 248BC, respectively.) The holography measurement was made at 11.451 GHz, using the linearly polarized beacon from geosynchronous orbit satellite ECS-1 as the illumination source. The satellite was positioned 13° east of Greenwich. At DSS 63, this satellite appears at elevation 40.2 degrees, azimuth 154.4 degrees. Its diurnal motion

was less than  $\pm 0.1$  degrees in elevation and azimuth at the time of the tests. The e.i.r.p. of the ECS-1 beacon is 11 dBW. With the receiver bandwidth of 2.5 Hz, this provided a beam peak signal-to-noise ratio of 73 dB. The panels of the 64m DSS 63 antenna typically measure about 3.5m X 1.5m. In order to determine the mean position and tilt of the panels, it is necessary for the resolution of the holography images to be at least as small as half the smallest panel dimension (Ref. 2). By recording a data array of 189 X 189, a resolution of 0.4m was achieved. The duration of the measurement was about 11 hours (Ref. 4).

The raw data, supplied to JPL by Eikontech, Ltd. was taken on 28 May 1985. The contractor corrected the recorded data for satellite motion, phase drift, pointing offset errors, baseline phase errors, and azimuth/elevation to rectangular u,v coordinate interpolation.

The digitized data array (far-field response) of 189 X 189 complex numbers (32 bits each for real and imaginary parts) was extended with zeros at the prescribed  $\Delta u$ ,  $\Delta v$  intervals to a 256 X 256 array and processed by the IFFT program. The result, the near-field aperture response of the antenna, was then further processed to display two effective maps: effective surface error map (derived from phase) and surface current map (amplitude). The complete processing and display cycle with no iteration took about 90 minutes. Before panel correction adjustment can be calculated, the data (of future "final" maps) must be fitted globally to the "best" parabola (a criteria that must be determined) and then locally best-fitted to determine panel screw point adjustments (finite element models including girders, beams, and panel skins are included in the software).

Damaged and locally misaligned panels can clearly be observed on the surface error map (especially on the color images). For example, the panel close to the lower right strut appears depressed (this panel suffered accidental damage during an earlier modification). Other grossly misaligned panels can clearly be observed.

Diffraction effects due to subreflector flange appear as exceptionally well-defined concentric rings close to the edge of the dish on both maps. Figures 3 and 4 show that independent computer simulation/analysis predictions of the aperture functions corresponding to Figures 1 and 2 agree well in describing the subreflector flange diffraction effect. The subreflector flange was designed for optimum 2295-MHz beam shaping (Figure 5, from Ref. 5).

Panel grids, subreflector blockage, quadripod blockage, and projected shadows have been intentionally overlayed on the "surface error map," while only direct (plane wave) optical

shadows have been overlayed on the “surface current map.” Spherical wave-shadowed regions of the surface error map generally contain meaningless phase information (the shadowed panels cannot be set from these data) and are therefore suppressed. The same regions of the surface current map are displayed, as valuable information can be gained from the shadow details.

### III. Summary

A key benefit of the microwave holography technique is that the mathematically transformed data contain information about the antenna beam formation, allowing precise know-

ledge not only about the detailed panel alignments on a “local” scale, but also, importantly, about the subreflector alignment on a “global” scale.

Considerable further detailed work is required to achieve full data reduction, analysis, and manipulation, and improved display capability. However, the fundamentals are now in place and operating. Work is needed to produce a more complete analysis capability and toward better understanding via computer simulation of the various error types (random noise, ordered pointing errors, subreflector/horn misalignments) and examination of needs at 32 GHz. Following completion of that work, attention needs to shift toward data acquisition capability.

### Acknowledgments

The authors wish to thank Boris Seidel, who contracted and supervised the supplying of the raw data used to produce the holography maps in this report, and Dan Bathker for his numerous helpful technical discussions.

### References

1. Rahmat-Samii, Y., “Surface Diagnosis of Large Reflector Antennas using Microwave Holographic Metrology: An Iterative Approach,” *Radio Science*, Vol. 19, No. 5, pp. 1205–1217, Sept.–Oct., 1984.
2. Rahmat-Samii, Y., “Microwave Holography of Large Reflector Antennas: Simulation Algorithms,” *IEEE Transaction on AP*, Vol. AP-33, No. 11, pp. 1174–1203, November 1985.
3. Rochblatt, D. J., and Seidel, B. L., “DSN Microwave Antenna Holography,” *TDA Progress Report 42-76*, pp. 27–34, October–December, 1983.
4. Godwin, M. P., Schoessow, E. P., and Richards, P. J., “Final Report on Holographic Tests at S-band and K-band on the DSS-63 64-Meter Antenna,” prepared for the Jet Propulsion Laboratory under Contract No. 956984, February 1986.
5. Bathker, D. A., “Radio Frequency Performance of a 210-ft Ground Antenna, X-band,” *JPL Technical Report 32-1417*, Jet Propulsion Laboratory, December 15, 1969.

ORIGINAL PAGE IS  
OF POOR QUALITY

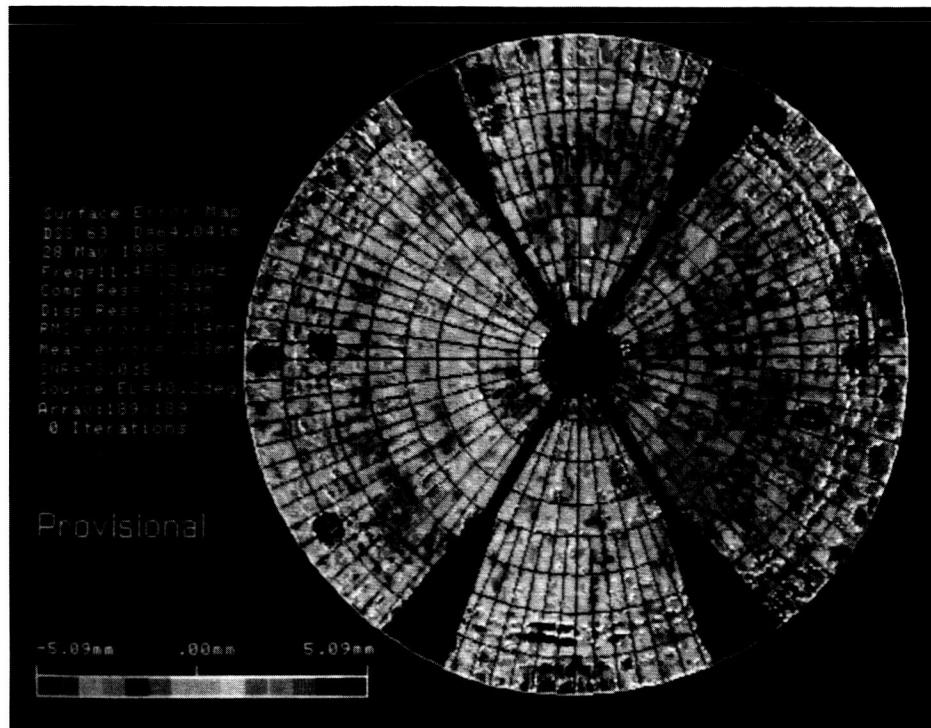


Fig. 1. DSS 63 64-m effective surface error map

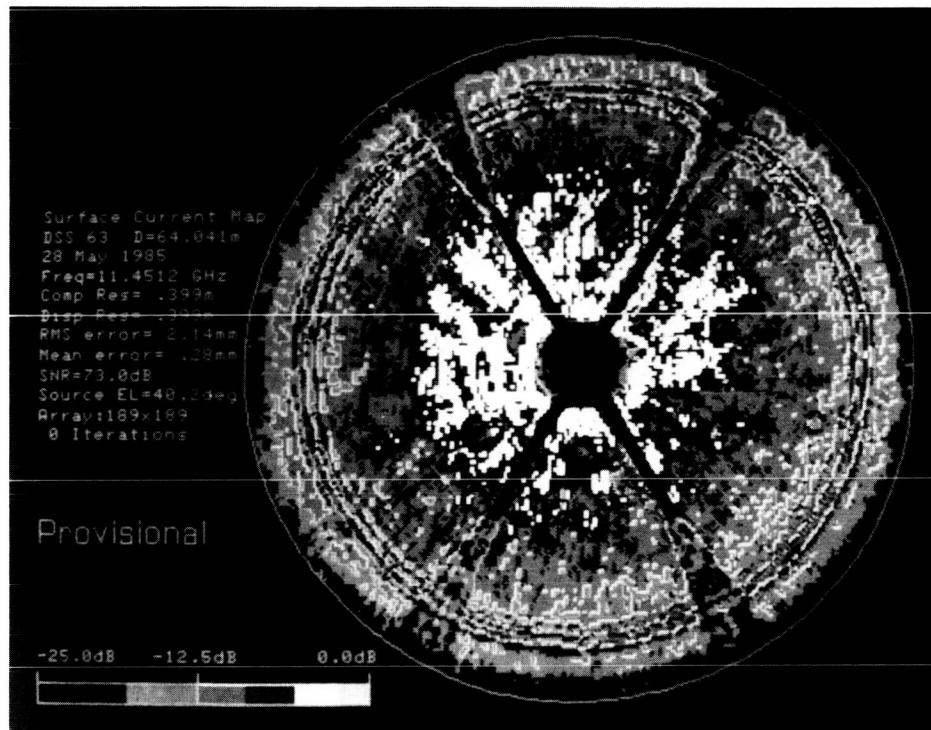
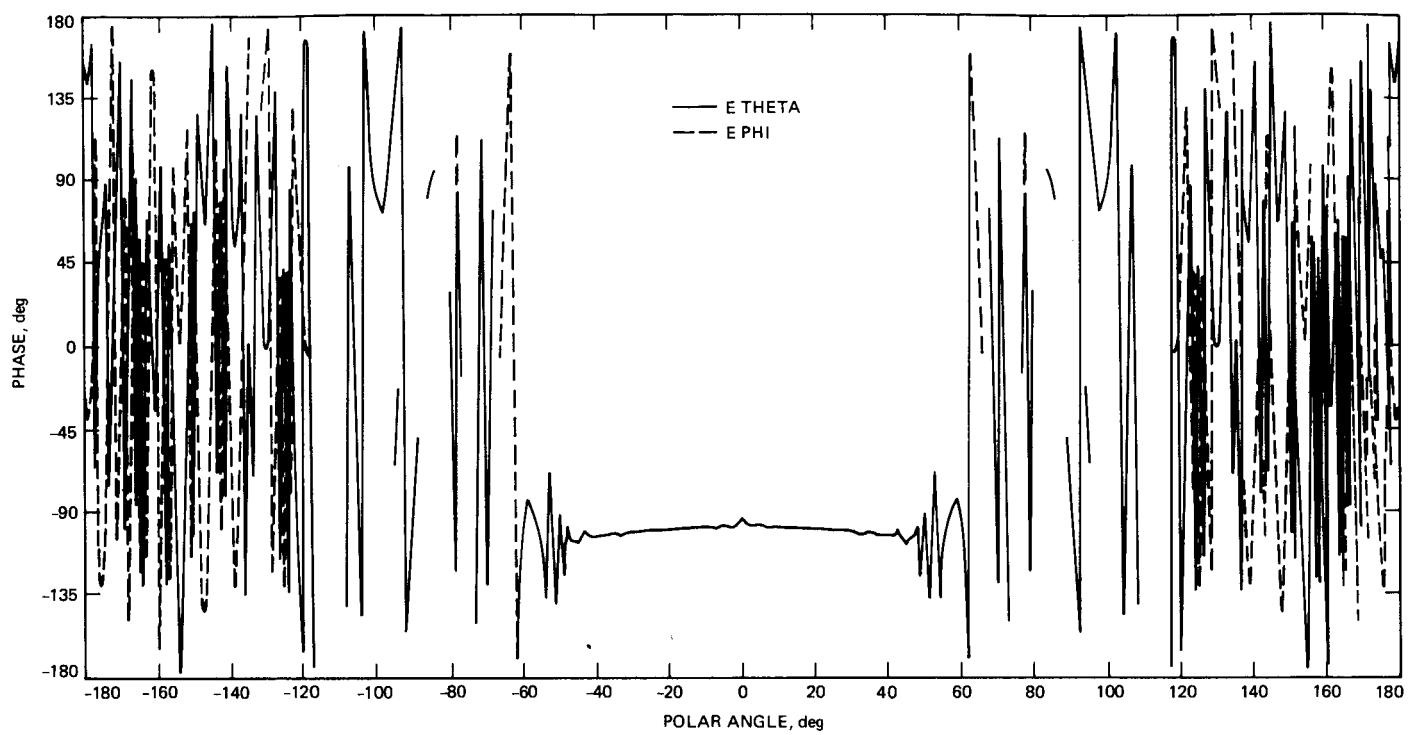
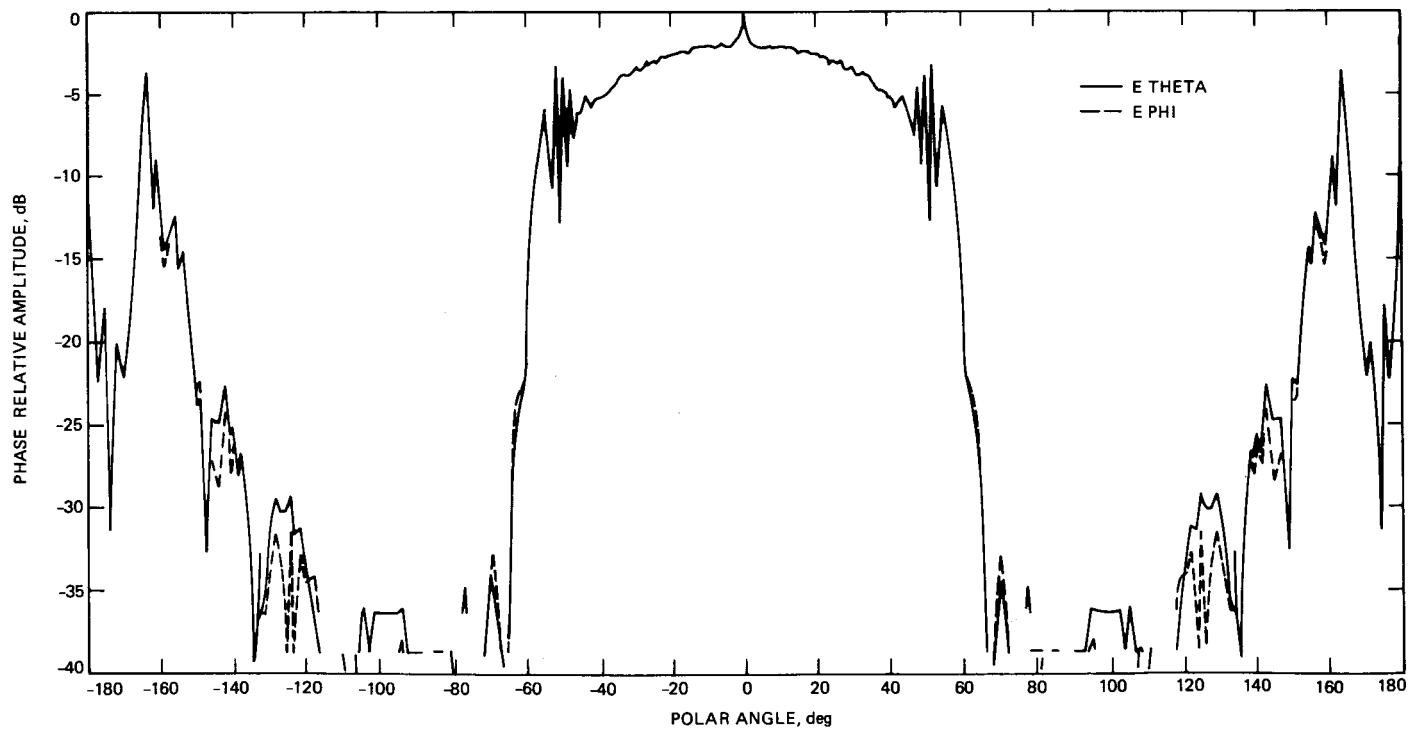


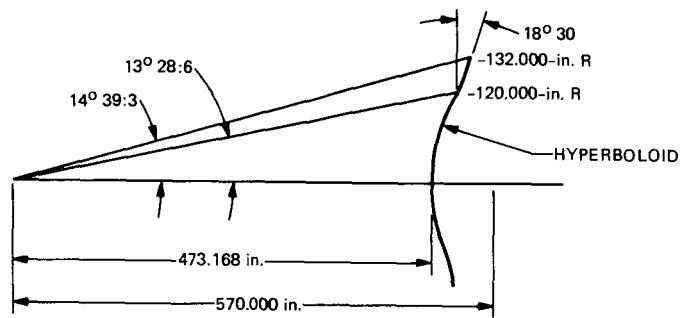
Fig. 2. DSS 63 64-m effective surface current map



**Fig. 3. Subreflector scattered radiation patterns, phase**



**Fig. 4. Subreflector scattered radiation patterns, amplitude**



**Fig. 5. 64-m subreflector configuration**

DII-33

Q-4

TDA Progress Report 42-87

N87-15340

July-September 1986

# A Single VLSI Chip for Computing Syndromes in the (255, 223) Reed-Solomon Decoder

I. S. Hsu, T. K. Truong, H. M. Shao, and L. J. Deutsch  
Communications Systems Research Section

*This article presents a description of a single VLSI chip for computing syndromes in the (255, 223) Reed-Solomon decoder. The architecture that leads to this single VLSI chip design makes use of the dual basis multiplication algorithm. The same architecture can be applied to design VLSI chips to compute various kinds of number theoretic transforms.*

## I. Introduction

Berlekamp (Ref. 1) has developed for JPL a bit-serial multiplication algorithm for the encoding of Reed-Solomon (RS) codes. This multiplication algorithm uses a dual basis over a Galois field. It is shown in Ref. 2 that by the use of this new dual basis multiplication algorithm, a (255, 223) Reed-Solomon encoder can be realized readily on a single VLSI chip with NMOS technology. Recently, based on the idea of Berlekamp, the authors (Ref. 3) developed a new dual basis multiplier which requires less transistors than normal basis (Ref. 4) or standard basis (Ref. 5) multipliers in VLSI implementation. This new dual basis multiplier can be used as a building block in the Reed-Solomon code decoder design (Ref. 6). It is demonstrated in this article how this dual basis multiplier is used as a basic cell to compute syndromes for a (255, 223) Reed-Solomon decoder (Ref. 7). The number of transistors needed in this design is about 6000. This is much less than it would be if the architecture used in the design of a 4-bit syndrome computing cell were adopted in this design.<sup>1</sup>

With that type of architecture, the expected number of transistors needed for the implementation of the syndrome computing chip is about 16,000. The major improvements that lead to the substantial reduction of the number of transistors needed are twofold:

- (1) The dual basis multiplication algorithm is used in the present design. In the previous design, the multiplier used is based on the normal basis multiplication algorithm developed by Massey and Omura (Ref. 8). This change results in a saving of about 200 transistors for each multiplier used in the syndrome computation. Totally, it is a reduction of  $200 \times 32 = 6400$  transistors for the syndrome computing chip. The number 32 comes from the fact that there are 32 syndrome subcells included in the (255, 223) Reed-Solomon decoder syndrome computer design.
- (2) In the previous design of the syndrome computing chip, the multipliers used are general-purpose multipliers. This means that both the multiplier and multiplicand are variable. This design did not take full advantage of the syndrome computation property. In

<sup>1</sup>H. M. Shao, "A VLSI Syndrome Computing Chip for Reed-Solomon Decoder," private communication, 1985.

computing the syndromes  $S_1, S_2, \dots, S_{32}$  of the received message, the multiplicand is always a fixed element. There is, hence, no need to use a general-purpose multiplier in syndrome computation. The general-purpose multiplier used in the previous design is replaced by a fixed multiplicand multiplier. One serial-to-parallel conversion unit is also saved by this modification. Since one serial-to-parallel conversion unit consists of two sets of 8-bit registers and several control gates, this saving means a reduction of about 100 transistors for each syndrome computing subcell. Alternatively, it is a saving of about  $32 \times 100 = 3200$  transistors for the whole chip.

In conclusion, the number of transistors saved for the present design over the previous design is  $6400 + 3200 = 9600$  transistors. That is, almost two thirds of the number of transistors are saved by these modifications.

## II. A VLSI Design for Computing Syndromes in a (255, 223) Reed-Solomon Decoder

In this section, a VLSI architecture developed in Ref. 7 is used to compute syndromes for a (255, 223) Reed-Solomon decoder over  $GF(2^8)$ . The dual basis multiplier developed in Ref. 3 is used to compute the inner product in the basic subcell of the syndrome chip. Figure 1 shows the overall diagram of the syndrome chip. In the figure, "Vdd" and "GND" are two power pins. "Phi-1" and "Phi-2" are the inputs for the two nonoverlapping clocks. "Start" is a signal that indicates the start of a received codeword. "Load" and "N-Load" are used to control the serial-to-parallel shifting of input data. These two signals are complements of each other. The shift operation is performed once for every eight bits. "Out" and "N-Out" are used to control the synchronization of input and output data. Again, these two are complementary signals. The function of these control signals will be explained in detail in the following section.

The operations that a syndrome computing chip performs are described below.

The syndromes of the received message for a (255, 223) Reed-Solomon code can be written as

$$S_k = \sum_{n=0}^{254} r_n \alpha^{nk}, \text{ for } 1 \leq k \leq 32 \quad (1)$$

where  $\{r_n\}$  is a received message. There are 255 symbols in a code word of a (255, 223) RS code. Hence,  $n$  ranges from 0 to 254 for computing the syndrome of each code word.

In order to develop a VLSI architecture to compute the  $S_k$ 's, first let the expression in (1) be rewritten in recursive form as

$$S_k = (\dots((0 + r_{254}) \cdot \alpha^k + r_{253}) \cdot \alpha^k \dots) + r_0, 1 \leq k \leq 32 \quad (2)$$

This expression is the famous Horner's rule. A VLSI architecture is developed to compute (2) as shown in Fig. 2. The function of each cell depicted in Fig. 2 can be best described by the register transfer relation

$$R_B \leftarrow R_B + R_A \times R_C$$

where " $\leftarrow$ " denotes the operation "is replaced by,"  $R_A$ ,  $R_B$ , and  $R_C$  are contents of registers A, B, and C, respectively. The received messages are sent into the chip bit-by-bit serially. They are in the order  $r_{254}, r_{253}, \dots, r_0$  for each received code word. Each incoming bit is then sent to all the subcells simultaneously. Figure 3 shows the block diagram of the syndrome computing chip. This chip consists of 32 subcells. Each subcell is responsible for computing a specific syndrome, i.e., subcell  $S_1$  computes syndrome  $S_1$ , subcell  $S_2$  computes syndrome  $S_2$ , etc. These subcells are arranged in the form of a 4 by 8 matrix with each element in the matrix a syndrome computing subcell. By this organization, the syndrome computing unit can be integrated with the polynomial expansion unit which accepts data from the syndrome computing unit (Ref. 6) without any matching problem. Each subcell shown in Fig. 3 performs the operation described in (2). The block diagram of a subcell is depicted in Fig. 4. As shown in Fig. 4, a syndrome computing subcell is divided into three units: (1) serial-to-parallel conversion unit, (2) multiplication unit, and (3) accumulator register unit. The detailed operation of each unit is explained in the following.

### A. Serial-to-Parallel Unit

This unit performs the serial-to-parallel operation. This unit is needed because, as mentioned above, the received messages are a stream of bits, one bit after the other, while the multiplication algorithm is designed to perform operations on the basis of symbols. That is, 8 bits are used to do the calculation simultaneously. The serial-to-parallel conversion is controlled by the signals "Load" and "N-Load." The signal "Load" is high for every 8 bits, which corresponds to the number of bits in a symbol in (255, 223) RS code. When signal "Load" is high, the serial-to-parallel operation is carried out. Otherwise, all the data are shifted forward bit-by-bit.

### B. Multiplication Unit

This unit performs the multiplication of an input symbol by a fixed finite element. Since the multiplicand is fixed, it

can be built in the circuit as a stream of exclusive-OR gates (Ref. 3). For example, in subcell  $S_1$ , the fixed multiplicand is  $\alpha^1$ ; therefore, only the output data from the second register in the serial-to-parallel unit is needed as shown in Fig. 4. In subcell  $S_2$ , the fixed multiplicand is  $\alpha^2$ , only the output data from the third register in the serial-to-parallel unit is required, etc.

### C. Accumulator Register Unit

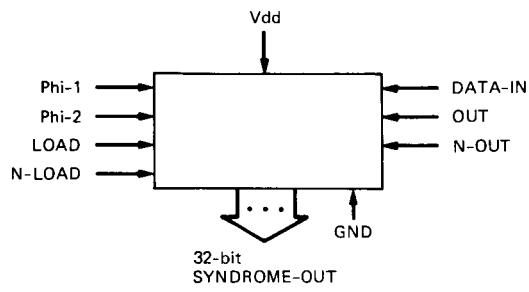
This unit is the temporary storage of the accumulated products for the operation described in Subsection A, above. At the 255th clock cycle for each input code word, the accumulated data in the registers are all shifted out bit-by-bit and added to the input message  $r_{254}$ . The result, which is the syndrome for this code word, is then shifted out bit-by-bit from each syndrome subcell. Syndromes  $S_1, S_2, \dots, S_{32}$  are shifted out in parallel. The data shift operations are controlled by signals "Out" and "N-Out." Initially, the control signal "N-Out" is high for 254 symbol times, i.e., it is high for message symbols  $r_{254}, r_{253}, \dots, r_1$ . During this period, the received messages are allowed to come into the chip and perform the operation as described by Eq. (2). The temporary sum for each syndrome subcell is stored in their respective accumulator register. When the 255th received symbol  $r_0$

comes in, the control signal "N-Out" is switched to low, and its complementary signal "N-Out" is changed to high. The result in the accumulator register in each syndrome subcell is then added to the incoming symbol  $r_0$  and the results, which are the calculated syndromes, are shifted out of the chip in parallel. This is due to the fact that the next step in RS code decoding process is to perform polynomial multiplication and to multiply the syndrome polynomial with the erasure polynomial. The polynomial multiplication unit is designed to accept data in parallel.

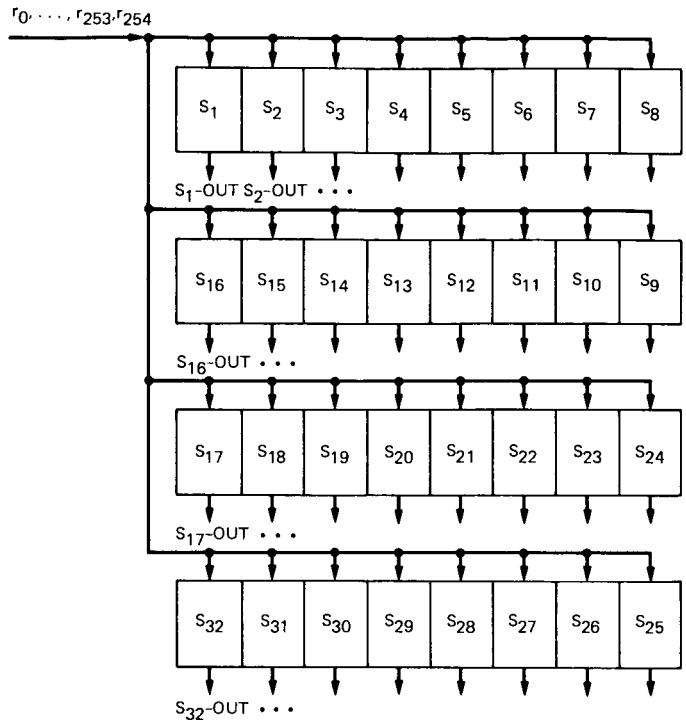
The layout of this syndrome chip is complete and is shown in Fig. 5. The layout of the whole chip was checked by the layout rule checker against a set of layout rules offered by the manufacturer. Logic, circuit, and timing simulations were performed thoroughly so that the risk of design errors is reduced to the least possible. This chip was sent to MOSIS (Ref. 9) for fabrication. After the chips are returned, they will be tested and evaluated. The technology used to fabricate the chip is 3  $\mu\text{m}$  NMOS. The operating frequency is estimated to be around 10 MHz. The power consumption of this chip is estimated to be 150 mW. The total number of transistors in this chip is approximately 6000, and the area of this chip is about  $6330 \times 2600 \mu\text{m}^2$ .

## References

1. Berlekamp, E. R., "Bit-Serial Reed-Solomon Encoders," *IEEE Trans. on Information Theory*, Vol. IT-28, No. 6, pp. 869-874, Nov. 1982.
2. Hsu, I. S., Reed, I. S., Truong, T. K., Wang, K., Yeh, C. S., and Deutsch, L. J., "The VLSI Implementation of a Reed-Solomon Encoder Using Berlekamp's Bit-Serial Multiplier Algorithm," *IEEE Trans. on Computers*, Vol. C-33, No. 10, pp. 906-911, Oct. 1984.
3. Hsu, I. S., Truong, T. K., Shao, H. M., Deutsch, L. J., and Reed, I. S., "A Comparison of VLSI Architecture of Finite Field Multipliers Using Dual, Normal or Standard Bases," presented at Fourth International Workshop on VLSI in Communications, Quebec, Canada, June 1986.
4. Wang, C. C., Truong, T. K., Shao, H. M., Deutsch, L. J., Omura, J. K., and Reed, I. S., "VLSI Architecture for Computing Multiplication and Inverses in  $GF(2^m)$ ," *IEEE Trans. on Computers*, Vol. C-34, No. 8, pp. 52-64, Aug. 1985.
5. Scott, P. A., Tarvares, S. E., and Peppard, L. E., "A Fast Multiplier for  $GF(2^m)$ ," submitted to *IEEE Trans. on Computers*.
6. Hsu, I. S., Shao, H. M., and Deutsch, L. J., "A Design Plan for a Single Chip Reed-Solomon Decoder," to appear in *TDA Progress Report*, Jet Propulsion Laboratory, Pasadena, CA.
7. Shao, H. M., Truong, T. K., Deutsch, L. J., Yuen, J. H., and Reed, I. S., "A VLSI Design of a Pipeline Reed-Solomon Decoder," *IEEE Trans. on Computers*, Vol. C-34, No. 5, pp. 393-403, May 1985.
8. Hsu, I. S., Deutsch, L. J., Truong, T. K., and Shao, H. M., "A VLSI Single Chip 8-Bit Finite Field Multiplier," *TDA Progress Report 42-83*, pp. 45-50, Jet Propulsion Laboratory, Pasadena, CA, Nov. 15, 1985.
9. The MOSIS Project, *The MOSIS System (What it is and how to use it)*, Publication ISI/TM-84-128, Information Science Institute, University of Southern California, Marina Del Rey, CA, 1980.

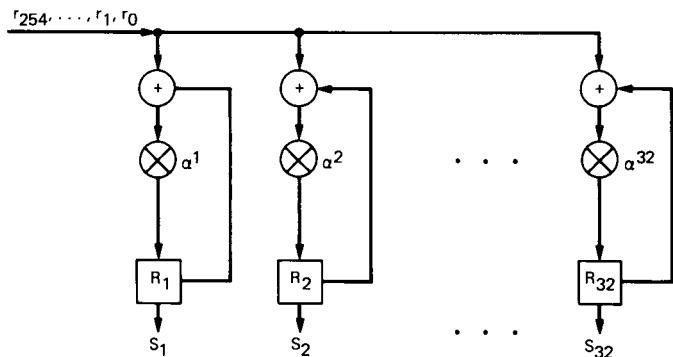


**Fig. 1. Overall diagram of the (255, 223) Reed–Solomon code syndrome computing chip**



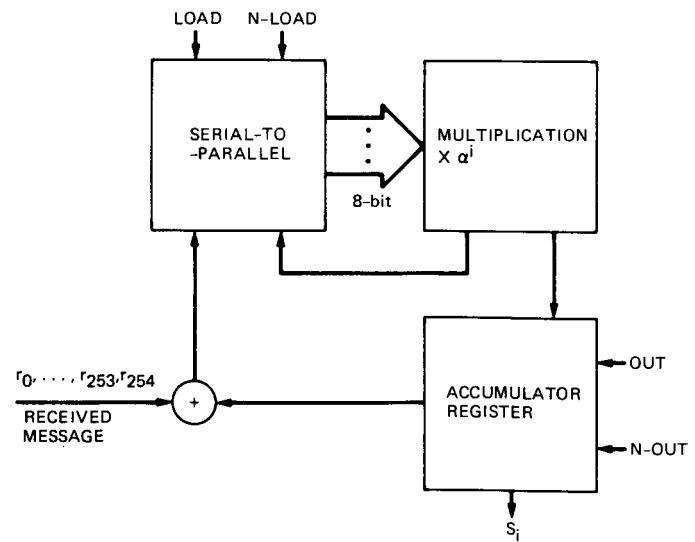
$S_i$ -OUT: THE COMPUTED  $i$ -th SYNDROME OUTPUT

**Fig. 3. Geometric position of each syndrome subcell in the syndrome computing chip**



$R_i$ :  $i$ -th 8-bit REGISTER

**Fig. 2. VLSI architecture for syndrome computing chip of the (255, 223) Reed–Solomon code decoder**



$S_i$ : COMPUTED  $i$ -th SYNDROME

**Fig. 4. Block diagram of the  $i$ -th syndrome computing subcell for computing syndrome  $S_i$**

ORIGINAL PAGE IS  
OF POOR QUALITY

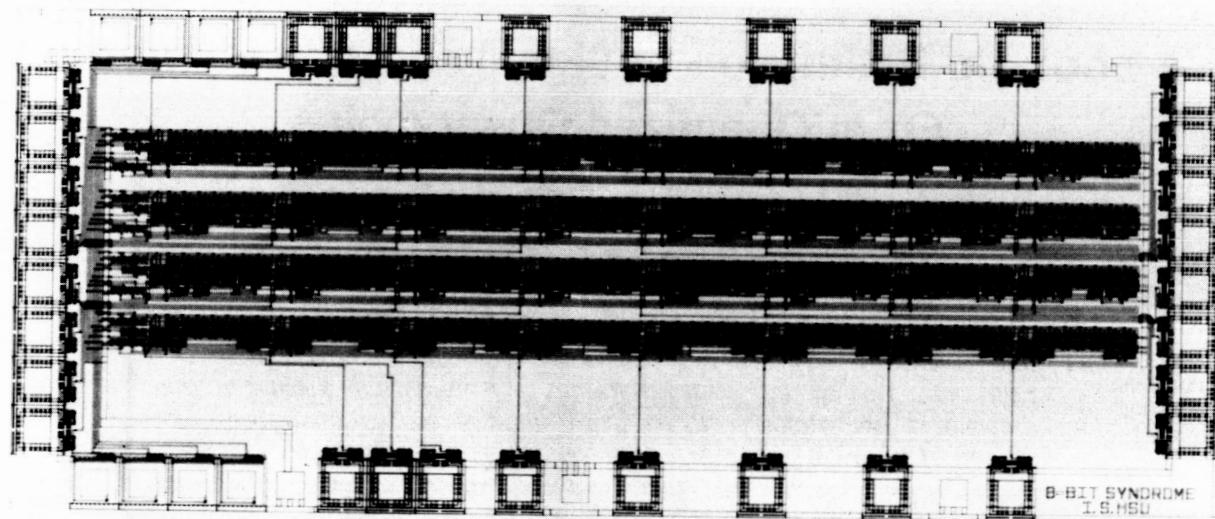


Fig. 5. Layout of the syndrome computing chip for the (255, 223) Reed-Solomon code decoder

4/2 - 32

10

# Table Look-Up Estimation of Signal and Noise Parameters From Quantized Observables

V. A. Vilnrotter and E. R. Rodermich  
Communications Systems Research Section

*In this article we examine a table look-up algorithm for estimating underlying signal and noise parameters from quantized observables. A general mathematical model is developed, and a look-up table designed specifically for estimating parameters from four-bit quantized data is described. Estimator performance is evaluated both analytically and by means of numerical simulation, and an example is provided to illustrate the use of the look-up table for estimating signal-to-noise ratios commonly encountered in Voyager-type data.*

## I. Introduction

In this article, we consider the problem of estimating signal and noise parameters from quantized samples of an observed waveform, by means of a table look-up algorithm. The waveform consists of binary antipodal signal plus additive Gaussian noise. The justification for the look-up table approach hinges on the observation that if the parameters of interest are single-valued functions of some computable quantities associated with the samples, then it should be possible to construct a table whose entries at the appropriate coordinates are the desired parameter estimates. Indeed, we shall show that for the class of problems under consideration, the first two absolute moments of the quantized samples can serve as entry coordinates to a two-dimensional look-up table. Thus, the problem reduces to that of finding accurate estimates for the first two absolute moments. Since accurate estimates of the required absolute moments can usually be found (as long as a large number of independent samples are available), the table look-up algorithm may often be employed to obtain quick and accurate parameter estimates.

The above approach was originally suggested to the authors by W. J. Hurd of the Communications Systems Research Section and was used in non-real-time symbol-stream combining of Voyager telemetry data (Ref. 1). In the following sections, suitable mathematical models are developed and employed to evaluate estimator performance. An example is provided to illustrate the use of the look-up table for estimating signal-to-noise ratios in data streams typical of actual Voyager data.

## II. Estimator Model

The table look-up algorithm may be used to estimate signal amplitude and noise standard deviation from a sequence of random variables obtained by synchronously integrating the received noise-corrupted waveform over the duration of each binary symbol. Thus, if the received waveform is expressed as

$$r(t) = D(t)(A/T) + n(t) \quad (1)$$

(where  $A$  is the integrated symbol amplitude and  $n(t)$  is additive Gaussian noise) then synchronous  $T$ -second integration over the  $i$ th symbol interval yields

$$r_i = \int_{(i-1)T}^{iT} r(t) dt = D_i A + n_i \quad (2)$$

where  $D_i$  is the integral of the antipodal modulation  $D(t)$  (hence  $D_i$  takes on the value  $\pm 1$  with probability  $P_1$  and  $P_{-1}$ , respectively) and the noise samples  $n_i$  are assumed to be independent, zero-mean Gaussian random variables with variance  $\sigma^2$ . The probability density of each sample can be found by averaging the conditional densities (conditioned on the data) over the a priori statistics of  $D_i$ . Assuming stationary statistics we can suppress the subscript “ $i$ ” and write

$$\begin{aligned} p_r(R) &= P_1 p_r(R | D = 1) + P_{-1} p_r(R | D = -1) \\ &= (2\pi\sigma^2)^{-1/2} \left\{ P_1 e^{-(R-A)^2/2\sigma^2} + P_{-1} e^{-(R+A)^2/2\sigma^2} \right\} \end{aligned} \quad (3)$$

Consider the case where the time samples are subjected to  $L$ -bit quantization prior to processing. In particular, let the quantized samples  $q$  take on integer values in the range

$$-2^{(L-1)} \leq q \leq 2^{(L-1)} - 1 \quad (4)$$

and let the probability that  $q$  takes on the integer value  $j$  be denoted  $p_j$  for any  $i$ :

$$p_j = \begin{cases} \int_j^{j+1} p_r(R) dR; & -2^{(L-1)} + 1 \leq j \leq 2^{(L-1)} - 2 \\ \int_{2^{(L-1)}-1}^{\infty} p_r(R) dR; & j = -2^{(L-1)} \text{ or } 2^{(L-1)} - 1 \end{cases} \quad (5)$$

In typical applications the a priori probabilities of the data symbols are equal ( $P_1 = P_{-1}$ ), in which case the quantization probabilities may be expressed as

$$p_j = \begin{cases} G(j) - G(j+1); & -2^{(L-1)} + 1 \leq j \leq 2^{(L-1)} - 2 \\ G(j); & j = -2^{(L-1)} \text{ or } 2^{(L-1)} - 1 \end{cases} \quad (6a)$$

where

$$G(j) = \frac{1}{4} \left\{ \operatorname{Erfc} \left( \frac{j-A}{\sqrt{2}\sigma} \right) + \operatorname{Erfc} \left( \frac{j+A}{\sqrt{2}\sigma} \right) \right\} \quad (6b)$$

and

$$\operatorname{Erfc}(\gamma) = \frac{2}{\sqrt{\pi}} \int_{\gamma}^{\infty} e^{-y^2} dy \quad (6c)$$

It is desired to obtain estimates of the parameters  $A$  and  $\sigma$  (denoted  $\hat{A}$  and  $\hat{\sigma}$ ) from the sample absolute mean  $\theta_1$  and sample mean-square  $\theta_2$ , which are defined as

$$\theta_1 = \frac{1}{N} \sum_{i=1}^N |q_i| \quad (7a)$$

$$\theta_2 = \frac{1}{N} \sum_{i=1}^N q_i^2 \quad (7b)$$

where  $N$  is the number of samples observed. Since the quantized observables are random variables, the absolute moment vector  $\theta \triangleq (\theta_1, \theta_2)$  is a random vector whose statistics depend on the number of quantization intervals and on the total number of samples as well as on the signal and noise parameters embedded in the received waveform.

Under the assumption that a large number of independent samples are used, the joint distribution of  $\theta$  can be determined by means of the multivariate central limit theorem (Refs. 2, 3). Defining the components of the auxiliary random vector  $x$  as

$$x_1 = |q| - \bar{|q|} \quad (8a)$$

$$x_2 = q^2 - \bar{q}^2 \quad (8b)$$

and forming the sum

$$z = \frac{1}{\sqrt{N}} \sum_{i=1}^N x_i \quad (9)$$

it follows that as  $N$  grows without bound, the distribution function of  $z$  approaches that of a jointly Gaussian random vector with covariance matrix

$$\Lambda_z = E[z^T z] = E[x^T x] = \Lambda_x \quad (10)$$

where, for any  $i$ , the components of  $\Lambda_x$  are

$$\lambda_{11} = E(x_1^2) = \bar{q}^2 - (\bar{|q|})^2 \quad (11a)$$

$$\lambda_{12} = \lambda_{21} = E(x_1 x_2) = \bar{|q|^3} - \bar{|q|} \bar{q}^2 \quad (11b)$$

$$\lambda_{22} = E(x_2^2) = \bar{q}^4 - (\bar{q}^2)^2 \quad (11c)$$

Here, as well as in equations (8), the overbar denotes expectation. The  $k$ th absolute moment is defined as

$$\bar{|q|^k} = \sum_j |j|^k p_j \quad (12)$$

Since the absolute moment vector can be expressed as

$$\theta = \frac{1}{\sqrt{N}} z + \bar{\theta} \quad (13a)$$

$$\bar{\theta} = (\bar{|q|}, \bar{q}^2) \quad (13b)$$

The central limit theorem implies that for sufficiently large  $N$ ,  $\theta$  may be modeled as a Gaussian random vector with mean value vector  $\bar{\theta}$  and covariance matrix

$$\Lambda_\theta = \frac{1}{N} \Lambda_x \quad (14)$$

Observe that as  $N$  approaches infinity, each component of  $\Lambda_\theta$  approaches zero, implying that for any choice of input parameters and quantization levels  $\theta$  approaches the associated mean value vector  $\bar{\theta}$ . It is this property of the absolute moment vector that originally motivated the construction of a table look-up algorithm for estimating signal and noise parameters.

A look-up table may be constructed in the following manner. Given  $L$ , the mean value vector  $\bar{\theta}$  may be determined for any choice of  $(A, \sigma)$ , using equation (12), and a dense matrix generated for each component of  $\bar{\theta}$  over the desired range of  $A$  and  $\sigma$ . Numerical interpolation then enables one to compute estimates of  $A$  and  $\sigma$  corresponding to a uniformly spaced grid over  $(\bar{\theta}_1, \bar{\theta}_2)$ .

In the current application, a look-up table was constructed for obtaining the estimates  $(\hat{A}, \hat{\sigma})$  from four-bit quantized samples ( $L = 4$ ), over the range  $0.5 \leq (A, \sigma) \leq 5.5$ . It was useful to introduce the predistortion transformation  $(\theta_1, \theta_2) \rightarrow (\bar{\theta}_1, \Theta)$  where

$$\Theta = [\max(0., 1.39211 - 0.0519 \theta_1 - (\sqrt{\theta_2}/\theta_1))]^{1/2} \quad (15)$$

in order to improve resolution over certain regions of interest. A graphical representation of the resulting look-up table is shown in Fig. 1, illustrating the functional dependence of the parameter estimates on the transformed coordinates  $(\theta_1, \Theta)$ . Note that some regions in the  $(\theta_1, \Theta)$  plane fall outside the range of the look-up table. In general, points within these external regions may be assigned special values (i.e. negative values) to distinguish them from valid parameter estimates. The user may then invoke a different estimator, or take other appropriate action, when one of these special values is encountered. In our case, the estimates depend on whether  $\Theta$  is less than or greater than 0.3:

$$\Theta > 0.3; \begin{cases} \hat{A} = \theta_1 \\ \hat{\sigma} = 0.5 \end{cases} \quad (16a)$$

$$\Theta \leq 0.3; \begin{cases} \hat{A} = 0.4 \\ \hat{\sigma} = \theta_1 \sqrt{\pi/2} \end{cases} \quad (16b)$$

Given measured values of  $\theta_1$  and  $\Theta$  falling within the range of the look-up table, the table entry coordinates are the integers  $I\theta_1$  and  $I\Theta$  most nearly satisfying

$$\theta_1 = 0.05(I\theta_1 + 19) \quad (17a)$$

$$\Theta = 0.005(I\Theta - 1) \quad (17b)$$

with  $1 \leq I\Theta \leq 95$  and  $1 \leq I\theta_1 \leq 81$ . All table entries for  $A$  and  $\sigma$  were converted to integers in the range 0 to 255. Designating the integer table entries for  $A$  and  $\sigma$  as  $IA$  and  $I\sigma$ , respectively, final estimates are obtained using

$$\hat{A} = 0.4 + 0.02 IA(I\theta_1, I\Theta) \quad (18a)$$

$$\hat{\sigma} = 0.4 + 0.02 I\sigma(I\theta_1, I\Theta) \quad (18b)$$

Since the table entries are integers, resolution is limited to 0.02 for both estimates. Examples of the effects of this "granularity" will be provided in the next section.

### III. Estimator Performance

The performance of the table look-up estimator has been evaluated both analytically and by means of numerical simula-

tion. The details of the analysis are presented in the Appendix. We shall consider the analytical results first.

### A. Performance Analysis

It was shown previously that the absolute sample moment vector can be decomposed into the sum of a deterministic and a zero-mean Gaussian component, provided the total number of observed samples is sufficiently great. For convenience, represent the sample moment components as

$$\theta_1 = \bar{\theta}_1 + \alpha_1 \quad (19a)$$

$$\theta_2 = \bar{\theta}_2 + \alpha_2 \quad (19b)$$

where the mean values are defined in Eq. (13b), and  $(\alpha_1, \alpha_2)$  are small Gaussian random variables of mean zero and covariance matrix  $\Lambda_\alpha = \Lambda_\theta$  (see Eq. [14]). The estimates  $(\hat{A}, \hat{\sigma})$  are obtained by inverting the formulas for  $(\theta_1, \theta_2)$  in terms of  $A$  and  $\sigma$ . Assuming that the inverse equations can be linearized over small enough regions in the  $(\theta_1, \theta_2)$  plane, the estimates may also be approximated as the sum of deterministic and zero-mean Gaussian components

$$\hat{A} \approx A + \epsilon_1 \quad (20a)$$

$$\hat{\sigma} \approx \sigma + \epsilon_2 \quad (20b)$$

Therefore,  $\epsilon_1$  and  $\epsilon_2$  represent random estimation errors resulting from the effective additive noise components  $(\alpha_1, \alpha_2)$ . Due to the Gaussian assumption, the error vector is completely characterized (in the statistical sense) by its covariance matrix  $\Lambda_\epsilon$ , with components

$$\epsilon_{11} = E(\hat{A} - A)^2 \quad (21a)$$

$$\epsilon_{12} = \epsilon_{21} = E(\hat{A} - A)(\hat{\sigma} - \sigma) \quad (21b)$$

$$\epsilon_{22} = E(\hat{\sigma} - \sigma)^2 \quad (21c)$$

In order to evaluate this matrix, observe that under the linearizing approximation (which is valid for large  $N$ ) the transpose of the effective noise vector can be written as

$$\alpha^T = C \epsilon^T \quad (22)$$

where  $C$  is the Jacobian of the forward transformation:

$$C = \begin{bmatrix} \frac{\partial \theta_1}{\partial A} & \frac{\partial \theta_1}{\partial \sigma} \\ \frac{\partial \theta_2}{\partial A} & \frac{\partial \theta_2}{\partial \sigma} \end{bmatrix} \quad (23)$$

An explicit calculation of the components of the Jacobian is performed in the Appendix, assuming four-bit quantization. With  $C^{-1}$  the inverse of  $C$ , the estimation error vector becomes

$$\epsilon^T = C^{-1} \alpha^T \quad (24)$$

while its covariance matrix may be represented as

$$\Lambda_\epsilon = E(\epsilon^T \epsilon) = C^{-1} \Lambda_\alpha (C^{-1})^T \quad (25)$$

This covariance matrix can be evaluated for any choice of  $A$ ,  $\sigma$ ,  $L$  and  $N$  using equations (11), (14) and (A-5). The error covariance matrix was determined at the four internal points designated in Fig. 1, namely at coordinates (2,2), (3,2), (4,2) and (3,3) in the  $(A, \sigma)$  domain. The results are displayed in Table 1. The expectations in Table 1 refer either to ensemble averages (obtained from the linearized analysis) or to sample averages (obtained from the simulation described in Section III.B).

### B. Performance Simulation

The performance of the estimator can also be evaluated by means of numerical simulation. This can be accomplished by generating random sequences with the appropriate statistics, performing the quantization operation, taking the magnitude and square of each sample, and adding up the desired number of terms to obtain a simulated sample absolute moment vector. Next, the look-up table is entered at the coordinates specified by the components of the simulated vector, and estimates of the desired parameters are obtained. The appropriate sample statistics may be determined by repeating the above procedure a large number of times, using independent random sequences each time. However, if each random sequence consists of a large number of terms, then the multivariate central limit theorem may be invoked to generate samples of the absolute moment vector directly with the proper statistics. This latter approach was adopted in the current application.

The purpose of the simulation is to characterize estimator performance at a given value of signal level  $A$ , noise standard deviation  $\sigma$ , quantization level  $L$  and sample size  $N$ . Once  $A$  and  $\sigma$  are specified, one computes  $\bar{\theta}_1 = |q|$ ,  $\bar{\theta}_2 = q^2$  and simulates the noise vector  $\alpha$  by means of the central limit theorem approximation. The simulated value of  $\alpha$ , denoted  $\tilde{\alpha}$ , can be represented as a zero-mean, unit variance Gaussian random vector  $u$  (with independent components) scaled by a matrix  $\Gamma$  as

$$\tilde{\alpha} = u\Gamma \quad (26)$$

where

$$\Gamma = \begin{bmatrix} \sqrt{\lambda_{11}} & \lambda_{12}/\sqrt{\lambda_{11}} \\ 0 & \sqrt{\lambda_{22} - \frac{\lambda_{12}^2}{\lambda_{11}}} \end{bmatrix} \frac{1}{\sqrt{N}} \quad (27)$$

It is readily verified that

$$E(\tilde{\alpha}^T \tilde{\alpha}) = \Lambda_{\alpha} = \Lambda_{\theta} \quad (28)$$

hence, in the limit of large  $N$ , the simulated approximation  $\tilde{\alpha}$  has the same statistics as the actual noise vector  $\alpha$ . The simulated sample absolute moment vector  $\tilde{\theta}$  is obtained using

$$\tilde{\theta} = \bar{\theta} + \tilde{\alpha} \quad (29)$$

Next, the look-up table is entered at the coordinates specified by  $\tilde{\theta}$ , and the estimates  $(\hat{A}, \hat{\sigma})$  recorded. This procedure is repeated  $K$  times, after which sample statistics are computed from the entries of the resulting  $(2 \times K)$  array. The value  $K = 10^4$  was chosen for the simulation. The distribution of the parameter estimates are displayed in Fig. 2 for typical simulation runs, using  $N = 10^3$  in each case. These "scatter diagrams" serve to reveal the structure of the look-up table grid, illustrate noise-induced scattering, and indicate correlation in the estimates near the input coordinates. However, it is not possible to form an accurate mental picture of either the mean or the "spread" of the resulting sample distributions, since the number of times each visible grid point occurred is not evident. Therefore, sample statistics were computed at the four designated coordinates to provide this additional information. The simulation statistics are displayed in Table 1, directly under the analytical results. When computing these sample statistics, points that were out of range were ignored, in order to remove dependence on the choice of the ad hoc estimator employed. Most simulation runs did not register out of range points, but even when such points occurred, only a few were recorded per simulation run.

Analysis and simulation results were found to be in good agreement. Errors in the sample means were negligible, indicating that the estimates are unbiased. Discrepancies in covariance are attributed to deviation from the linear model. Indeed, by letting  $N = 10^4$  at the point  $(3, 3)$ , the agreement between simulation and analysis improved, as shown in the last two rows of Table 1.

### C. An Application

Finally, the look-up table was employed to estimate symbol signal-to-noise ratio (SNR) for Voyager data typical of that used in the symbol stream combiner. Here symbol SNR is defined as

$$\text{SNR} = A^2/2 \sigma^2 \quad (30)$$

while its estimate is obtained directly from  $\hat{A}$  and  $\hat{\sigma}$  as

$$\widehat{\text{SNR}} = \hat{A}^2/2 \hat{\sigma}^2 \quad (31)$$

The estimator bias  $B$ , defined as

$$B = E(\widehat{\text{SNR}}) - \text{SNR} \quad (32)$$

is a measure of the difference between the mean of the estimate and the actual SNR. For an unbiased estimator,  $B = 0$ . Normalizing by the true SNR, one may define the ratio

$$R = E(\widehat{\text{SNR}})/\text{SNR} = 1 + \frac{B}{\text{SNR}} \quad (33)$$

which, when expressed in decibels, becomes

$$R (\text{dB}) = 10 \log_{10} [E(\widehat{\text{SNR}})/\text{SNR}] \quad (34)$$

(Thus, if  $R (\text{dB}) = 3$ ,  $E(\text{SNR}) = 2 \text{ SNR}$ , whereas if  $R (\text{dB}) = -3$ ,  $E(\text{SNR}) = 0.5 \text{ SNR}$ ). In addition, if the bias is small compared to the true SNR, then Eq. (34) reduces to

$$R (\text{dB}) \approx 10 \log_{10} (e) [B/\text{SNR}] \quad (35)$$

providing a convenient measure of the fractional bias.

The quantity  $R (\text{dB})$  was determined by simulation for true SNRs ranging from  $-1$  to  $+1$  dB, and various signal levels characteristic of Voyager data. (In the simulation, expectation is approximated by the sample mean.) The number of samples used for determining each estimate was  $N = 5700$  in agreement with the value used for the near-real-time symbol-stream combiner. The results, shown in Fig. 3, confirm that SNR estimates are virtually unbiased over the indicated range. The sample standard deviation of estimation error was also computed for

each signal level, and found to be nearly constant, approximately equal to 0.018 over the same range. These results indicate that the table look-up estimator examined in this article can be used to obtain unbiased estimates of symbol SNR over the range of interest for deep-space reception.

#### IV. Summary and Conclusions

A look-up table was constructed for the purpose of estimating signal and noise parameters from quantized samples of a noise-corrupted antipodal signal. Since the table is entered at coordinates derived from the first two sample absolute moments of the quantized observables, substantial errors in moment estimates lead directly to errors in the corresponding

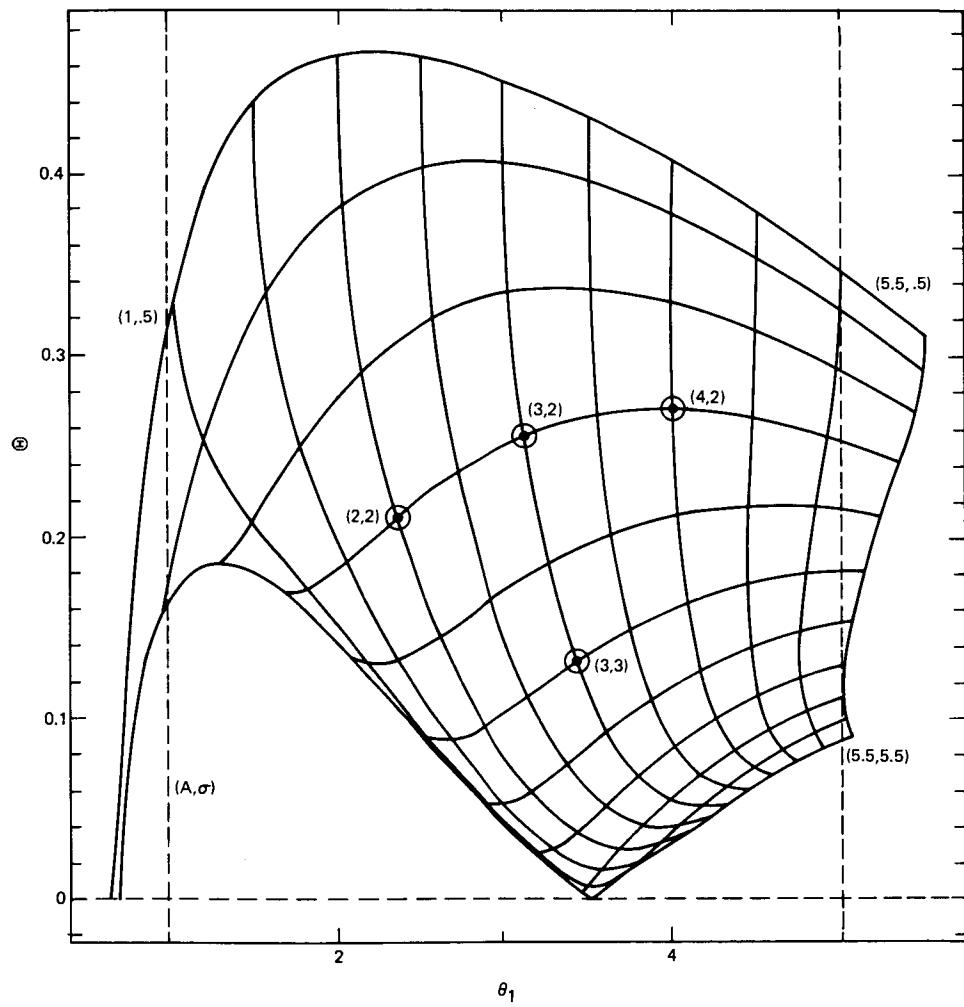
parameter estimates. Hence, the look-up table performs best when a large number of independent quantized samples are available to allow accurate determination of the required sample statistics. It should be observed, however, that perfect moment estimates do not necessarily yield perfect parameter estimates, due to the limited resolution inherent in any finite element table. The parameter estimates were found to be unbiased over the central region of the table, with random components that depend both on the sample size and on the spectral level of the additive noise process. Thus, the covariance matrix of the estimation error is a strong function of the coordinates at which the table is entered. It was also demonstrated that unbiased estimates of symbol SNR can be obtained, over a range of values and sample sizes characteristics of those encountered in processing actual Voyager data.

#### References

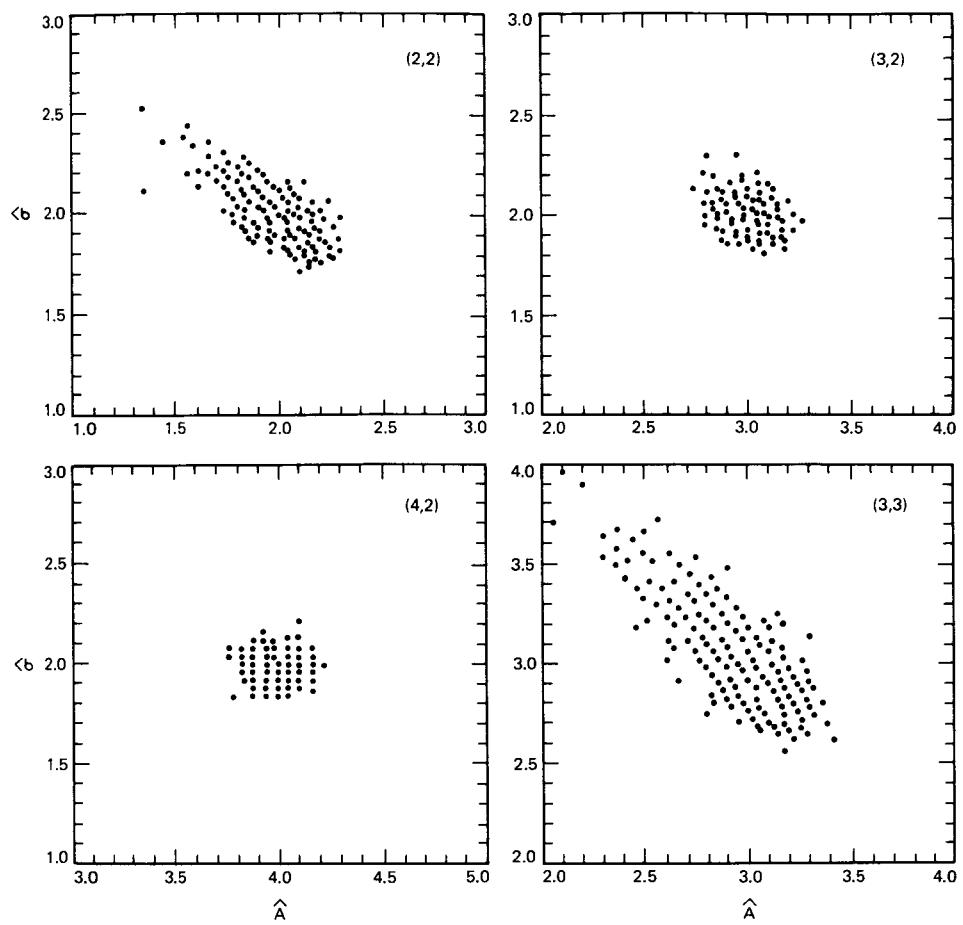
1. Hurd, W. J., Rabkin, J., Russell, M. D., Siev, B., Cooper, H. W., Anderson, T. O., and Winter, P. U., "Antenna Arraying of Voyager Telemetry Signals by Symbol Stream Combining," *TDA Progress Report 42-86*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 131-142, Aug. 15, 1986.
2. Wozencraft, J. M., and Jacobs, I. M., *Principles of Communication Engineering*, Chapt. 3., John Wiley & Sons, New York, 1965.
3. Wilks, S. S., *Mathematical Statistics*, Chapt. 9, John Wiley & Sons, New York, 1962.

**Table 1. Comparison of analysis and simulation results**

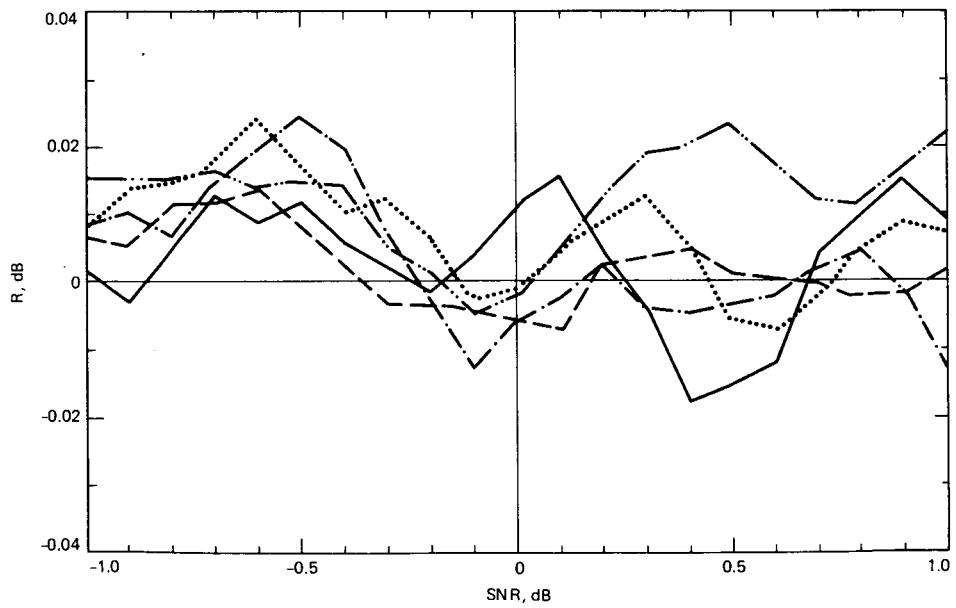
	$E(\hat{A})$	$E(\hat{\sigma})$	$E(\hat{A}\bar{\hat{A}})^2$	$E(\hat{A}\bar{\hat{A}})(\hat{\sigma}-\bar{\hat{\sigma}})$	$E(\hat{\sigma}\bar{\hat{\sigma}})^2$
$N = 10^3$					
Analysis	2.0	2.0	$1.258 \times 10^{-2}$	$-8.370 \times 10^{-3}$	$1.089 \times 10^{-2}$
Simulation	1.988	2.008	$1.578 \times 10^{-2}$	$-9.995 \times 10^{-3}$	$1.206 \times 10^{-2}$
Analysis	3.0	2.0	$5.242 \times 10^{-3}$	$-1.390 \times 10^{-3}$	$4.783 \times 10^{-3}$
Simulation	2.998	2.002	$5.606 \times 10^{-3}$	$-1.428 \times 10^{-3}$	$5.159 \times 10^{-3}$
Analysis	4.0	2.0	$4.535 \times 10^{-3}$	$-1.640 \times 10^{-4}$	$3.562 \times 10^{-3}$
Simulation	4.001	1.997	$5.640 \times 10^{-3}$	$7.2 \times 10^{-5}$	$3.572 \times 10^{-3}$
Analysis	3.0	3.0	$2.889 \times 10^{-2}$	$-2.379 \times 10^{-2}$	$3.640 \times 10^{-2}$
Simulation	2.967	3.024	$4.344 \times 10^{-2}$	$-3.429 \times 10^{-2}$	$4.375 \times 10^{-2}$
$N = 10^4$					
Analysis	3.0	3.0	$2.889 \times 10^{-3}$	$-2.379 \times 10^{-3}$	$3.640 \times 10^{-3}$
Simulation	2.996	3.002	$3.651 \times 10^{-3}$	$-2.623 \times 10^{-3}$	$3.509 \times 10^{-3}$



**Fig. 1. Look-up table representation**



**Fig. 2. Scatter diagrams at coordinates  $(A, \sigma)$**



**Fig. 3. The quantity  $R$  (dB) as a function of the true SNR**

## Appendix

In this appendix, formulas for the components of the Jacobian matrix  $C$  are derived, assuming four-bit quantization ( $L = 4$ ). It is convenient to write the  $\ell$ th absolute moment of the quantized observable  $q$  as

$$\bar{\theta}_\ell = E(|q|^\ell) = \frac{1}{2} \sum_{j=-8}^7 (|j|^\ell + |j+1|^\ell) f_j \quad (\text{A-1})$$

where

$$f_j = \int_{a_j}^{a_{j+1}} e^{-u^2/2} du / \sqrt{2\pi} \quad (\text{A-2a})$$

$$a_j = (j + A)/\sigma \quad -7 \leq j \leq 7 \quad (\text{A-2b})$$

and  $a_8 = +\infty$ ,  $a_{-8} = -\infty$ . Letting

$$\begin{aligned} e_j &= e^{-a_j^2/2}, \quad -7 \leq j \leq 7 \\ e_8 &= e_{-8} = 0 \end{aligned} \quad (\text{A-3})$$

it follows that

$$\frac{\partial f_j}{\partial A} = (e_{j+1} - e_j) / \sqrt{2\pi} \sigma \quad (\text{A-4a})$$

$$\frac{\partial f_j}{\partial \sigma} = [(j + A)e_j - (j + A + 1)e_{j+1}] / \sqrt{2\pi} \sigma^2 \quad (\text{A-4B})$$

Therefore, the components of the Jacobian  $C$  can be computed using

$$c_{\ell 1} = \frac{\partial \bar{\theta}_\ell}{\partial A} = \sum_{j=-8}^7 \frac{1}{2} (|j|^\ell + |j+1|^\ell) \frac{\partial f_j}{\partial A} \quad (\text{A-5a})$$

$$c_{\ell 2} = \frac{\partial \bar{\theta}_\ell}{\partial \sigma} = \sum_{j=-8}^7 \frac{1}{2} (|j|^\ell + |j+1|^\ell) \frac{\partial f_j}{\partial \sigma} \quad (\text{A-5b})$$

Having determined the components of  $C$ ,  $C^{-1}$  can be computed and used in Eq. (25) to evaluate the covariance matrix of the estimation error.

3/3 - 33  
P 11

N 87 - 15342

## A Comparison of Methods for DPLL Loop Filter Design

S. Aguirre, W. J. Hurd, R. Kumar, and J. Statman

Communications Systems Research Section

*Four design methodologies for loop filters for a class of digital phase-locked loops (DPLLs) are presented. The first design maps an optimum analog filter into the digital domain; the second approach designs a filter that minimizes in discrete time a weighted combination of the variance of the phase error due to noise and the sum square of the deterministic phase error component; the third method uses Kalman filter estimation theory to design a filter composed of a least squares fading memory estimator and a predictor. The last design relies on classical control theory, including rules for the design of compensators. Linear analysis is used throughout the article to compare different designs, and includes stability, steady state performance and transient behavior of the loops. Design methodology is not critical when the loop update rate can be made high relative to loop bandwidth, as the performance approaches that of continuous time. For low update rates, however, the minimization method is significantly superior to the other methods.*

### I. Introduction

For many years phase-locked loops have been a cornerstone in phase coherent communication systems. Analog implementations dominated the scenario until the late sixties. Then, as a result of rapid advances in the field of microelectronics, digital systems offering a myriad of advantages began to replace their analog counterparts. The trend now is to implement and design phase-locked loops digitally.

Researchers in the field have explored several avenues for the implementation and design of digital phase-locked loops (DPLLs). An excellent survey of theoretical and experimental work accomplished in this area up to 1981 is presented by Lindsey and Chie (Ref. 1). The study, however, is not mature yet. The bulk of the research has concentrated on different ways to mechanize the loop phase detector, but very little

effort has been spent in the design of loop filters. The typical design implements simply the discrete version of loop filters that have been widely used in analog phase-locked loops (APLLs) (Refs. 2, 3, 4, 5). This approach has a drawback, because the design process does not take into account the computational delays inherently present in any sampled-data system. These delays cause wider loop bandwidths than calculated from the continuous time equations, larger steady state phase errors for given loop bandwidths and dynamics, and reduced overall loop stability.

The classical control theory approach to improve performance would be to insert a compensator into the system. A variety of compensators have been proposed, with lead, lag, and lead/lag being the most commonly used. Nagrath and Gopal (Ref. 6) comment that lead/lag compensators are not

very useful when open loop transfer function poles are present on the imaginary axis if dealing with continuous time systems (on the unit circle for digital systems). This occurs with DPLLs, and tuned (second order) compensators are required for substantial improvements. The usefulness of this approach is, however, questionable. With a similar effort, one may employ more refined techniques and even obtain superior loop filters optimum in some sense.

This lack of design methodology for loop filters has led to the recent works on the methods compared in this article. Here, we present a systematic comparison of several loop filters that have been proposed for a class of DPLLs wherein each sample of the phase detector output represents the average phase error since the previous sample. We evaluate four different methodologies for the design of such filters. These are as follows:

- (1) Impulse Invariance Transformation
- (2) Minimization Method
- (3) Estimation-Prediction
- (4) Classical Control Theory Approach

Throughout the article, linear models for the PLLs are used, and the analysis is based on Z-transform theory due to the discrete nature of the loop component building blocks.

The performance of each loop is measured in terms of gain margin for stability, steady state phase error for unmodeled dynamics, and transient distortion (sum of the squares of the deterministic phase errors at the sampling instants). These criteria are typical in linear system analysis to assess the "quality" of a control system. For example, in case of DPLL, the closed-loop gain of the linearized model depends directly on the received signal power level. Thus to maintain the loop stability under power level variations, it is desirable to have high gain margin for stability.

In order to establish a fair basis of comparison, a convenient parameter (common to all the loops) must first be chosen. Then, the behavior of the different loops can be measured against this parameter. We find it convenient to use the one-sided loop noise bandwidth as the common element, because the bandwidth appears explicitly in expressions that characterize the variance of the phase estimation error, regarded by communication systems engineers as one of the most important measures of PLL performance.

The organization of the article is as follows. In Section II, the basic DPLL configuration studied here is defined. In Section III, the four methodologies for determining loop parameters are summarized. In Section IV, transfer functions result-

ing from the four methods are presented and compared. Loop stability, transient and steady state responses are compared in Section V. Finally, some general conclusions are drawn.

## II. Basic DPLL Configuration

Several ideas have been suggested for the implementation of a DPLL, usually based on the mechanization of the phase detector. A notable example is the loop that attempts to track the zero crossings of the incoming signal; consequently, the sampling intervals are nonuniform (Ref. 1). In this article, however, uniform sampling is assumed. The main components of the general DPLL considered here are depicted in Fig. 1. The integrate-and-dump circuit integrates the phase detector output over  $T$  s, henceforth called the loop update time. The result is applied to the loop filter  $F(z)$ , which is typically implemented in software, to produce the control signal. This signal drives a hardware numerically controlled oscillator (NCO), whose output frequency is proportional to the control signal. The output phase of the NCO is fed back to become the reference phase input to the phase detector, to close the loop.

The input signal in Fig. 1 is  $A \cos(wt + \theta) + n(t)$  where  $n(t)$  is white noise with one-sided power spectral density  $N_0$  W/Hz. The output of the sampler is  $A\phi_k + n_k$  where  $\phi_k$  is the average phase error over the last sampling period, and  $n_k$  is a white noise sample with variance  $\sigma_n^2 = N_0/2T$ . An equivalent linearized model of the DPLL considered here is illustrated in Fig. 2. The generic transfer function  $KN_g(z)$  takes into account the mathematical representation of the NCO, the mechanization of the phase detector and a normalized computational delay  $g$ . This normalized transport lag is the fraction of loop update time interval from the time that each phase measurement is made until the NCO input is updated. The effects of the phase detector integrate-and-dump filter, the transport lag  $g$ , and the NCO are derived elsewhere (Refs. 7, 8) and can be included in a single transfer function  $N_g(z)$  given below.

$$N_g(z) = \frac{T[(1-g)^2 z^2 + (1+2g-2g^2)z + g^2]}{2z^2(z-1)} \quad (1)$$

The DPLL implementation is facilitated when  $g = 1$  corresponding to maximum possible delay. For this important example, Eq. (1) reduces to

$$N_1(z) = \frac{T(z+1)}{2(z-1)} \frac{1}{z^2} \quad (2)$$

This can be recognized as the cascade of two computational delays and the discrete version of the integral operation using the trapezoidal rule.

### III. Design of Loop Filters

In this section we provide a synopsis of the underlying principles behind each methodology. Further details may be found in the references.

#### A. Impulse Invariance Transformation (IIT)

The first design (Ref. 7) involves the impulse invariance transformation (Ref. 9) of an optimum analog filter into a digital filter. Under this transformation, perfect integrators are approximated by

$$\frac{1}{s} \sim \frac{Tz}{z - 1} \quad (3)$$

Other transformations such as the backward difference and the right-side rectangular rule also indicate Eq. (3) as their equivalent sampled-data transformation. In this manner, the equivalent sampled-data loop filter parameters can be conveniently expressed in terms of the corresponding analog loop filter parameters such as bandwidth and damping ratio.

An immediate advantage of this approach is that the theory of continuous time loops is very mature, and a wealth of useful knowledge has been accumulated. This information in many instances can be applied directly to the resultant digital loops. As one would expect, the continuous and discrete time theories are very close when the loop update rate is very high compared to loop bandwidth.

#### B. Minimization (MIN)

The minimization method (Ref. 10) applies optimal control theory concepts so as to arrive at a set of optimum digital filters for various input dynamics. The input to the loop is assumed to be the sum of a phase signal and white noise, the two being independent of each other. If  $\sigma$  represents the closed loop rms phase jitter due to additive white noise and if the sum square of the deterministic phase error is represented by  $D^2$ , the design criterion is to minimize

$$E = \sigma^2 + \lambda D^2 \quad (4)$$

where  $\lambda$  is a Lagrange multiplier. The design procedure is carried out as follows: For given input phase dynamics, fixed update rate and  $N_g(z)$ , find a causal filter  $F(z)$  that minimizes  $E$  in Eq. (4). The overall loop transfer function can be obtained as a function of  $\lambda$ . Then, for any desired loop noise bandwidth, the multiplier  $\lambda$  can be obtained.

An analogous optimum design procedure using Wiener filtering theory was presented by Jaffe and Rechtin (Ref. 11), leading to the loop filters that found widespread use in APLLs.

The same standard optimization technique was utilized later by Gupta (Ref. 12) to design a digital-analog loop.

#### C. Estimation-Prediction (E-P)

The third method (Ref. 13) borrows estimation theory concepts and is particularly attractive to Kalman filter users. The idea is illustrated in Fig. 3. Assume that the normalization is  $A = 1$ . Since  $N_g(z)$  is known and realizable in hardware, it can also be implemented in software. This is done, and the same value of  $\hat{\theta}_k$  which is subtracted in the phase detector is added back to the phase detector output in software, generating  $\theta_k + n_k$  at the input to  $S(z)$ . On the basis of this signal, the digital filter  $S(z)$  predicts the frequency correction signal for the NCO and its equivalent software realization. The filter transfer function  $S(z)$  is assumed to be of the following type

$$S(z) = \frac{C(z) D(z)}{z^N N_g(z)} \quad (5)$$

where  $N$  is the number of transport lags,  $C(z)$  is an estimator and  $D(z)$  is a predictor. The design procedure is divided into five different steps. The first step is to select a model for the received phase that includes known dynamics plus process and measurement noise. The second step is to specify  $N_g(z)$ . The next step is the selection of an estimator. For this example, a least squares fading memory estimator was selected. The estimator obtains a state estimate (phase, frequency and acceleration) applying an exponentially decaying weight to past data and ignoring the pure delay. The fourth step is to compensate for the delay by designing an appropriate predictor. The predictor uses the state estimate to predict the phase  $N$  delays ahead. The last step checks for loop realizability and stability.

#### D. Classical Control Theory Approach (CCT)

The fourth approach is a heuristic method (Ref. 8). Here, classical control theory concepts are applied, including rules for the design of compensators. The idea is to select a realizable transfer function of the form

$$F(z) = \frac{A(z)}{B(z)} \frac{z^2}{C(z)} \quad (6)$$

The quadratic term in  $z$  cancels the transport lags, and  $A(z)$ ,  $B(z)$  and  $C(z)$  are polynomials in  $z$  such that  $F(z)$  is a realizable filter transfer function. Based on repeated trials and errors, convenient locations of the roots of these polynomials are selected. This is done via root locus analysis.

Basically, the goal is to produce an overall stable loop with reasonable transient and steady state performance, using well-known rules for the construction of root locus plots. The selec-

tion of the pole and zero locations of the filter is informal and requires some design experience.

## IV. Loop Filter Transfer Functions

In this section we present a collection of results obtained with each of the design procedures outlined in the previous sections. For the sake of completeness, we provide their corresponding filter transfer functions. For space limitations, detailed derivations are omitted.

### A. Impulse Invariance Transformation (IIT)

The sampled-data filter has the form (Ref. 7)

$$F(z) = G_1 + \frac{G_2}{1 - z^{-1}} + \frac{G_3}{(1 - z^{-1})^2} \quad (7)$$

where

$$\left. \begin{aligned} G_1 &= rd/AKT \\ G_2 &= rd^2/AKT \\ G_3 &= krd^3/AKT \\ d &= \frac{4B_A T}{r} \left( \frac{r-k}{r-k+1} \right) \end{aligned} \right\} \quad (8)$$

In Eq. (8), the coefficients are those employed in continuous time loop filters. The parameter  $r$  is typically 2 or 4, and is equal to  $4\xi^2$  where  $\xi$  is the damping ratio. The parameter  $k$  is a type 3 loop gain component ( $k = 0$  for type 2 loop), with typical values ranging from  $1/4$  to  $1/2$ . The coefficient  $B_A$  is the noise bandwidth of the underlying analog loop. The parameters  $A$  and  $K$  are those appearing in Fig. 2.

### B. Minimization (MIN)

For a type 2 system the loop filter is given by (Ref. 9)

$$F(z) = \frac{2}{KT} \times \frac{(h_0 z - h_1) z^2}{\{4az^2 + (8a + 4b)z + (5a + 3b + c - d)\} (z - 1)} \quad (9)$$

The coefficients  $a, b, c$ , and  $d$  satisfy the identity

$$az^3 + bz^2 + cz + d = (z + 1)(\bar{a}z^2 + \bar{b}z + \bar{c}) \quad (10)$$

where  $\bar{a}, \bar{b}$ , and  $\bar{c}$  satisfy the following set of equations

$$\left. \begin{aligned} \bar{a}\bar{c} &= 1 \\ \bar{a}\bar{b} + \bar{b}\bar{c} &= -4 \\ \bar{a}^2 + \bar{b}^2 + \bar{c}^2 &= 6 + r \\ r &= \frac{\lambda T}{\bar{N}_0} \\ \bar{N}_0 &= N_0/A^2 \end{aligned} \right\} \quad (11)$$

where  $N_0$  = one-sided spectral density of input noise. The previous set of equations produces cancellation of a zero at  $z = -1$  by a pole at  $z = -1$ . Imperfect cancellation can cause instability problems; therefore, a slight modification to the previous equations is used where the factor  $(z + 1)$  in Eq.(10) is replaced by  $(z + q)$ . The value of  $q$  may be in the range of 0.9 to 0.999.

The type 3 loop filter has the form

$$F(z) = \frac{2}{KT} \times \frac{\{\tilde{C}z^2 + (\tilde{B} - 2\tilde{C})z + (\tilde{A} + \tilde{C} - \tilde{B})\} z^2}{(z - 1)^2 \{16az^2 + 16(3a + b)z + [16(6a + 3b + c) - \tilde{C}]\}} \quad (12)$$

The various filter coefficients are given in terms of a set of four nonlinear simultaneous equations that will not be reproduced here. The set of equations again produces a pole-zero cancellation at  $z = -1$ , that demands a slight modification. The reader may refer to Ref. 10 for a simple method to obtain the values of these coefficients.

### C. Estimator-Predictor (E-P)

The loop filter is given by (Ref. 13)

$$F(z) = \frac{S(z)}{1 - N_g(z) S(z)} \quad (13)$$

with  $S(z)$  as indicated in Eq. (5). For a type 2 loop with a fading memory filter with estimator decay factor  $\alpha$ ,  $0 < \alpha < 1$ , the loop filter transfer function is given by

$$F(z) = \frac{(az - b)z^2}{c[z^2 + (1 + b)z + b](z - 1)} \quad (14)$$

where

$$\left. \begin{array}{l} a = 3 - 4\alpha \\ b = 2(1 - \alpha) \\ c = T/2 \end{array} \right\} \quad (15)$$

$$P_1 = -0.173$$

$$P_2 = -0.999$$

$$z_1 = z_2 = 0.960$$

The fading memory type 3 loop is

$$F(z) = \frac{(az^2 + bz + c)z^2}{d(z^2 + ez + c)(z - 1)^2} \quad (16)$$

where

$$\left. \begin{array}{l} a = 6 - 9\alpha \\ b = 9\alpha - 8 \\ c = 3(1 - \alpha) \\ d = T/2 \\ e = 4 - 3\alpha \end{array} \right\} \quad (17)$$

Special attention must be paid in the implementation of  $S(z)$ , because a pole-zero cancellation at  $z = -1$  occurs as in the minimization method.

#### D. Classical Control Theory (CCT)

The last design procedure assumes a normalized computational delay  $g = 1/2$  in Eq. (1), while the other three methods assume  $g = 1$ . This, however, complicates the loop implementation. For a type 2 loop, the filter used in Ref. 8 is

$$F(z) = \frac{(z - z_1)z^2}{(z - P_1)(z - P_2)(z - 1)} \quad (18)$$

where

$$P_1 = -0.173$$

$$P_2 = -0.999$$

$$z_1 = 0.960$$

For a type 3 loop,

$$F(z) = \frac{(z - z_1)(z - z_2)z^2}{(z - P_1)(z - P_2)(z - 1)^2} \quad (19)$$

where

$$P_1 = -0.173$$

$$P_2 = -0.999$$

$$z_1 = z_2 = 0.960$$

It is noted that these parameters were chosen for a specific application, in which the product of the one-sided loop noise bandwidth  $B_L$  times  $T$  is in the neighborhood of 0.15. No design rules are given for other  $B_L T$ .

## V. Performance Comparisons

An indication of the relative merits of the individual configurations can be acquired using standard techniques of linear control systems. In this section, we employ three common criteria: (1) computation of stability gain margin, (2) calculation of steady state errors, and (3) calculation of transient distortion, defined as the sum of the squared phase estimation errors at the sampling instants. The output in all cases is graphical. In the design of PLLs, it is customary to treat the one-sided noise bandwidth as the independent variable and characterize loop behavior as a function of this parameter. We find it convenient to use, instead, noise bandwidth normalized (multiplied) by the loop update time.

In computing the results, it should be noted that Ref. 10 used symbols  $B_L$  and  $B$  for unnormalized and normalized two-sided loop bandwidth respectively, whereas we use them in the more conventional way to denote one-sided loop bandwidth. For the CCT approach, the results illustrated for the design example (Ref. 8) assume an update time  $T = 1$  s.

#### A. Stability and Gain Margin

It is well known that second order, type 2 APPLs are unconditionally stable, and that third order, type 3 APPLs are unstable for low loop gains, but stable otherwise (Refs. 2, 3, 4, 5). Digital PLLs, however, are only conditionally stable. Type 2 DPLLs are unstable for high loop gains, and type 3 DPLLs are unstable for both, low and high loop gains. To differentiate these two extreme cases in computing gain margins, we use the terms "lower" and "upper" gain margin, respectively.

In Fig. 4 we examine the stability of the various configurations for a type 2 loop. The CCT loop is somewhat superior because it assumes a normalized computational delay  $g = 1/2$ . The other three loops assume  $g = 1$  and have performances that differ among themselves by 2 dB at most, with the MIN method being the best.

In Fig. 5 we plot the lower gain margin for type 3 loops. It is important to emphasize that the CCT loop was designed with fixed compensation, not depending on desired  $B_L T$ . With this compensation, it cannot attain normalized bandwidths smaller than 0.06, because the loop becomes unstable. Any redesign for lower  $B_L T$  would change both upper and lower gain margins. Upper gain margins for type 3 loops are depicted in Fig. 6. The MIN method is slightly better than the IIT and E-P methods in upper gain margin, and slightly worse in lower gain margin.

## B. Steady State Performance

For both APPLLs and DPLLs, the steady state phase error due to step acceleration and jerk is approximately proportional to the inverse of second and third power of the bandwidth for types 2 and 3 loops, respectively. It is convenient then, to define the phase error coefficients  $C_2$  for type 2 loops and  $C_3$  for type 3 loops. The use of these coefficients is advantageous because they exhibit slower variations with bandwidth than steady state errors. Thus, the steady state phase errors are related to bandwidth and phase error coefficients in the following way:

Type  $n$ :

$$\phi_{ss} = \theta^{(n)} \left( \frac{C_n}{B_L} \right)^n \quad (20)$$

where  $\theta^{(n)}$  denotes the  $n$ th highest nonzero derivative of the input phase.

Figure 7 concentrates on type 2 loops. Notice from Eq. (20) that a smaller phase error coefficient implies a smaller steady state phase error. The MIN method is uniformly better than all of the other methods for all  $B_L T$ . For normalized bandwidths smaller than 0.02, both the IIT and MIN filter are superior to the E-P method and produce essentially the same results. This is because the E-P method used a fading memory filter, which has more damping than the IIT filter with  $r = 2$ . For larger bandwidths, the MIN filter is better. The poorest performance is provided by the CCT filter.

Figure 8 presents similar results and conclusions for type 3 loops. Again, the MIN method is best for all  $B_L T$ .

The importance of these results is that there are cases when significant performance improvements can be realized using the MIN method. Suppose  $T$  cannot be reduced, due to implementation limitations. In such situations using the MIN method of design, for a specified maximum steady state phase

error, a smaller value of loop noise bandwidth is obtained compared to that achievable from other design methods.

For example, for the type 3 loop, using Fig. 8,  $C_3/B_L T$  is approximately 6.7 at  $B_L T = 0.2$  for the MIN method and at  $B_L T = 0.3$  for the IIT and E-P methods. Thus a 1.8 dB (factor of 1.5) higher loop SNR can be achieved for the same lag error using the MIN method. In some cases, for very wide bandwidths, only the MIN method is satisfactory.

## C. Transient Performance

The integral square error or transient distortion is defined here as the sum of the squares of the deterministic phase error components at the loop update instants. The summation runs from zero to infinity. The integral square error is plotted in Figs. 9 and 10 for type 2 and 3 loops. In the former case we applied a unit phase ramp, computed the square error and normalized (divided by  $T^2$ ) the result. Type 3 loops assumed a unit frequency ramp, and the result was normalized by  $T^4$ . The MIN loop is the best in both cases, as it should be due to the optimization procedure. The CCT loop is the worst for type 2 (except for large normalized bandwidths), and the E-P loop is the worst for type 3 loops. In this last case, the gap between the MIN and the E-P loops is several orders of magnitude. This is probably due to selecting fading memory filters for the E-P case.

## VI. Conclusions

Four classes of digital filters have been presented that have been proposed recently for a class of DPLLs wherein each sample of the phase detector output represents the average phase error since the previous sample. Whereas the filters in the first class are the mapped versions of the corresponding optimum analog filter, the filters in the second class are derived using optimal control theory. A suboptimal version of fading memory Kalman filter/predictor results in filters of class 3, while filters of the last class are designed on the basis of classical control theory.

On the basis of the results depicted in Figs. 4-10 we conclude that the filter obtained using the minimization method is the best in all regards, except in low gain margin for type 3 loops, where it is the worst. Since its lower gain margin is adequate, it is normally the best choice.

When the system is not update rate limited, so that  $B_L T$  is small, the Impulse Invariance or continuous time analogy is very good. In fact, the Impulse Invariance and minimization filter performances converge to the same values for all the criteria. This agrees with intuition, because in the limit as the normalized bandwidth goes to zero, the Impulse Invari-

ance filter converges to a continuous time filter whose transfer function is derived using the same optimization technique that is employed by the minimization filter.

The specific E-P loops considered here use fading memory filters with a damping greater than that of IIT loops with  $r = 2$ . Therefore, they do not have as good a performance as the IIT and MIN loops even for small  $B_L T$ .

The loops derived from classical control theory were designed only for a limited range of  $B_L T$ . They perform

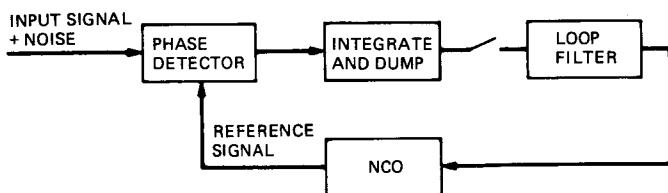
reasonably well there. Their main disadvantage is lack of an overall design method, resulting in loops with performance depending on the designer's skill, experience and intuition.

The main instance in which it is important to use the minimization method over the IIT method is when  $B_L T$  is not small. This occurs when the update rate has a maximum, due to implementation restrictions. Performance is then significantly better for the minimization method than for all other methods studied. In some cases, only the minimization method results in a usable design.

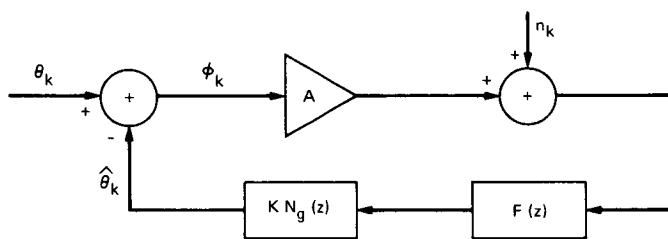
## References

1. Lindsey, W. C., and Chie, C. M., "A Survey of Digital Phase-Locked Loops," *Proceedings of the IEEE* 69, No. 4, pp. 410-431, April 1981.
2. Lindsey, W. C., *Synchronization Systems in Communication and Control*. Englewood Cliffs, N.J., Prentice-Hall, 1972.
3. Lindsey, W. C., and Simon, M. K., *Telecommunication Systems Engineering*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1973.
4. Gardner, F. M. *Phaselock Techniques*, 2nd Ed., Wiley, New York, 1979.
5. Holmes, J. K., *Coherent Spread Spectrum Systems*, John Wiley and Sons Inc., New York, N.Y., 1982.
6. Nagrath, I. J., and Gopal, M., *Control Systems Engineering*, Halsted Press, 1982.
7. Aguirre, S., and Hurd, W. J., "Design and Performance of Sampled Data Loops for Subcarrier and Carrier Tracking," *TDA Progress Report 42-79*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 81-94, Nov. 15, 1984.
8. Simon, M. K., and Mileant A., "Digital Filters for Digital Phase-Locked Loops," *TDA Progress Report 42-81*, Jet Propulsion Laboratory, Pasadena, California, pp. 81-93, April 15, 1985.
9. Oppenheim, A. V., and Schafer, R. W., *Digital Signal Processing*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1975.
10. Kumar, R., and Hurd W. J., "A Class of Optimum Digital Phase Locked Loops for the DSN Advanced Receiver," *TDA Progress Report 42-83*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 63-80, Nov. 15, 1985; also to appear in the *Proceedings of the 25th IEEE Conference on Decision and Control*, Athens, Greece, Dec. 1986.
11. Jaffe, R., and Rechtin, E., "Design and Performance of Phase-Lock Circuits Capable of Near-Optimum Performance Over a Wide Range of Input Signal and Noise Levels," *IRE Trans. Inform. Theory IT-1*, pp. 66-76, Mar. 1955.
12. Gupta, S. C., "On Optimum Digital Phase-Locked Loops," *IEEE Trans. Commun. Technol. COM-16*, pp. 340-344, Apr. 1968.

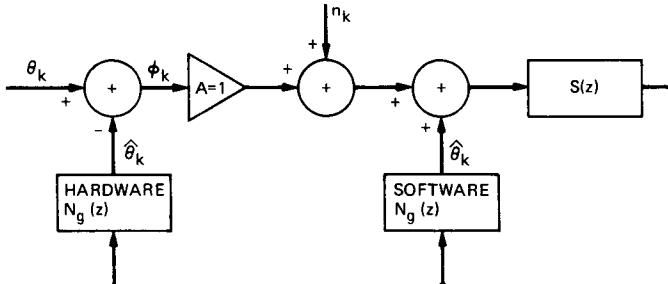
13. Statman, J. I., and Hurd, W. J., "An Estimator-Predictor Approach to PLL Loop Filter Design," *TDA Progress Report 42-86*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 77-89, Aug. 15, 1986.
14. Jury, E. I., *Theory and Application of the Z Transform Method*, New York, John Wiley and Sons, 1964.



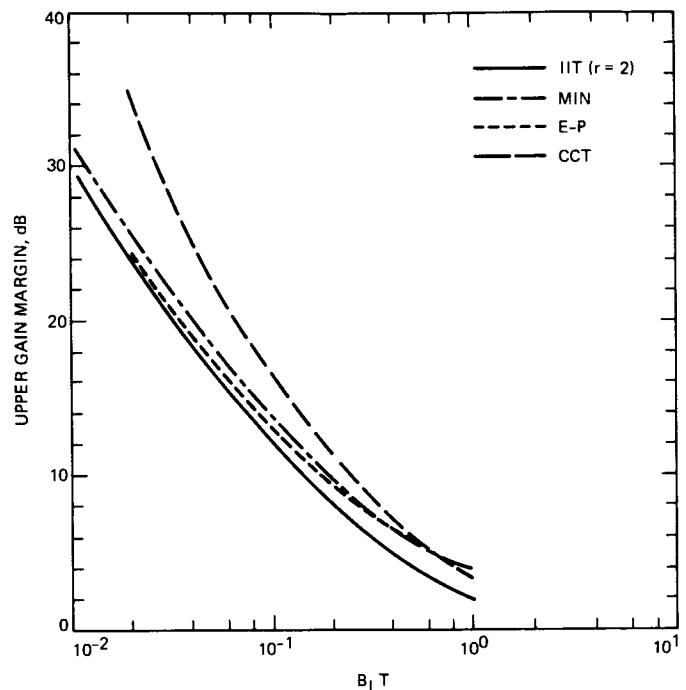
**Fig. 1. Digital phase-locked loop**



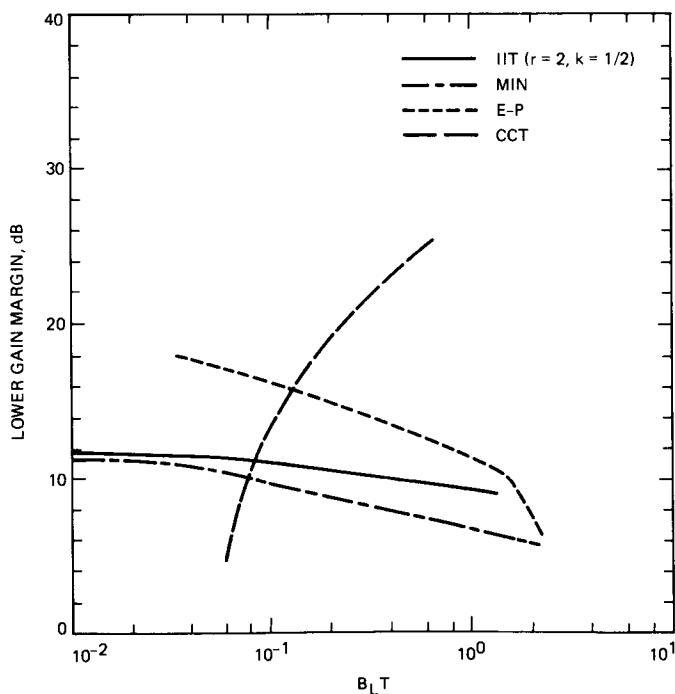
**Fig. 2. Linear baseband sampled data loop model**



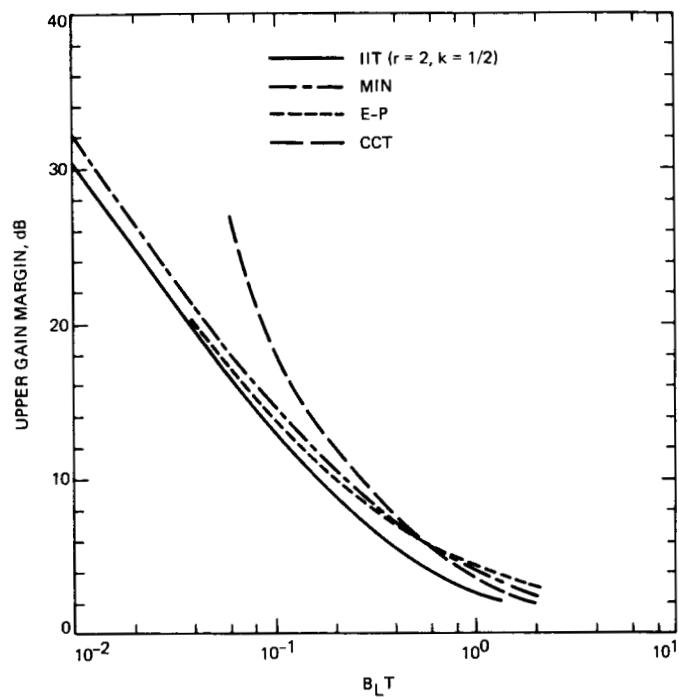
**Fig. 3. DPLL with estimator-predictor**



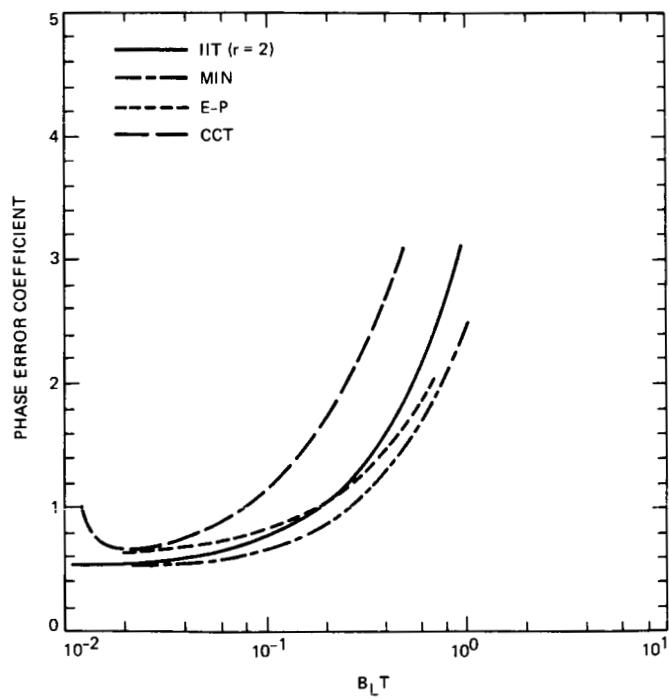
**Fig. 4. Upper gain margin for type 2 DPLL**



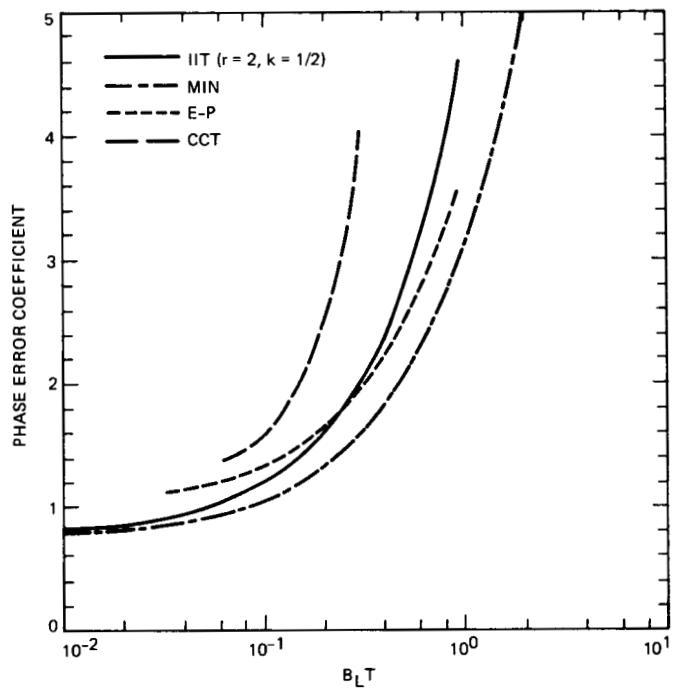
**Fig. 5. Lower gain margin for type 3 DPLL**



**Fig. 6. Upper gain margin for type 3 DPLL**



**Fig. 7. Steady state phase error performance of type 2 DPLL**



**Fig. 8. Steady state phase error performance of type 3 DPLL**

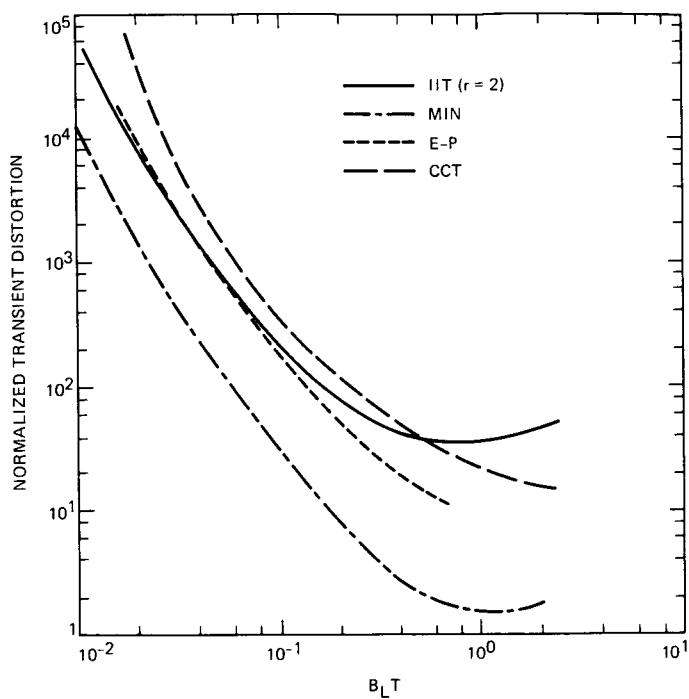


Fig. 9. Transient response of type 2 DPLL

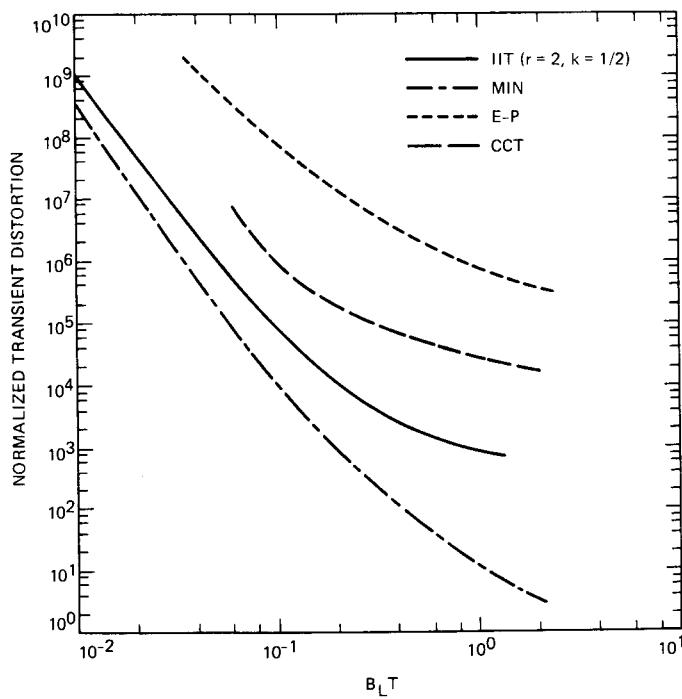


Fig. 10. Transient response of type 3 DPLL

# A Generalized Algorithm to Design Finite Field Normal Basis Multipliers

C. C. Wang

Communications Systems Research Section

*Finite field arithmetic logic is central in the implementation of some error-correcting coders and some cryptographic devices. There is a need for good multiplication algorithms which can be easily realized. Massey and Omura recently developed a new multiplication algorithm for finite fields based on a normal basis representation. Using the normal basis representation, the design of the finite field multiplier is simple and regular. The fundamental design of the Massey-Omura multiplier is based on a design of a product function. In this article, a generalized algorithm to locate a normal basis in a field is first presented. Using this normal basis, an algorithm to construct the product function is then developed. This design does not depend on particular characteristics of the generator polynomial of the field.*

## I. Introduction

The finite field  $GF(2^m)$  is a number system containing  $2^m$  elements. Its attractiveness in practical applications stems from the fact that each element can be represented by  $m$  binary digits. The practical application of error-correcting codes makes considerable use of computation in  $GF(2^m)$ . Both the encoding and decoding devices for the important Reed-Solomon codes must perform computations in  $GF(2^m)$  (Refs. 1, 2). The decoding device for the binary BCH codes also must perform computation in  $GF(2^m)$  (Refs. 1, 2). On the other hand, recent advances in secret communications, such as encryption and decryption of digital messages, also require the use of computation in  $GF(2^m)$  (Ref. 3). Hence, there is a need for good algorithms for doing multiplication in a finite field.

Yeh, Reed and Truong (Ref. 4) presented a design for performing multiplication in  $GF(2^m)$  which is suitable for VLSI implementation. In their design, the elements in the field are represented by a canonical basis  $\{1, \alpha, \alpha^2, \alpha^3, \dots, \alpha^{m-1}\}$  where  $\alpha$  is a root of an irreducible polynomial of degree  $m$  over  $GF(2)$ . Some other previous work on multipliers in  $GF(2^m)$  by Bartee and Schneider (Ref. 5), Gallager (Ref. 6), and Laws and Rushforth (Ref. 7) is also based on the canonical basis of  $GF(2^m)$ . However, these circuits are not as well suited for use in VLSI systems, due to irregular wire routing and complicated control problems as well as non-modular structure or lack of concurrency (Ref. 8).

Recently, Massey and Omura (Ref. 9) invented a multiplier which obtains the product of two elements in the finite field  $GF(2^m)$ . In their invention, they utilize a normal basis

of the form  $\{\alpha, \alpha^2, \alpha^4, \dots, \alpha^{2^{m-1}}\}$  to represent elements of the field. In this basis, again, each element in the field  $GF(2^m)$  can be represented by  $m$  binary digits.

In the normal-basis representation the squaring of an element in  $GF(2^m)$  is readily shown (Ref. 10) to be a simple cyclic shift of its binary digits. In the normal basis representations, multiplication requires the same logic function for any one bit of the product as it does for any other (Ref. 10). The generation of adjacent product digits differs only in the inputs, which are cyclically shifted versions of one another, to this product function. Hence, designing a Massey-Omura multiplier is exactly the same as designing a product function. In Ref. 10, a pipeline architecture suitable for VLSI design has been developed for a Massey-Omura multiplier of  $GF(2^m)$ . In comparison with the multiplier designed in Ref. 4, the Massey-Omura multiplier is much simpler.

In Ref. 10, the design of a Massey-Omura multiplier is based on a normal basis

$$\left\{ \alpha, \alpha^2, \alpha^{2^2}, \dots, \alpha^{2^{m-1}} \right\} \quad (1)$$

which is the set of roots of an irreducible polynomial

$$P(x) = x^m + c_1 x^{m-1} + \dots + c_m \quad (2)$$

over  $GF(2^m)$ .

In general, to verify the linear independence of the roots in Eq. (1) is difficult. A straightforward way to do this is to represent  $\alpha^{2^i}$ ,  $i = 0, 1, \dots, m - 1$ , by  $m$ -dimensional vectors in canonical basis  $\{1, \alpha, \alpha^2, \dots, \alpha^{m-1}\}$  and then to check whether the  $m \times m$  matrix composed by the above  $m$  vectors is nonsingular. For large  $m$ , this method requires a great number of computations. Peterson and Weldon (Ref. 2) list a set of irreducible polynomials of degree  $m \leq 34$  over  $GF(2)$  for which the roots are linearly independent.

For the case of  $m = 2^n$ , Perlis (Ref. 11) has shown that a necessary and sufficient condition for the above set (Eq. [1]) to be a normal basis of  $GF(2^m)$  is the trace of  $\alpha$  obeying the relation

$$Tr(\alpha) \triangleq \alpha + \alpha^2 + \alpha^{2^2} + \dots + \alpha^{2^{m-1}} = 1$$

or, equivalently, the coefficient  $c_1$  in Eq. (2) is 1. He also gave necessary and sufficient conditions for a normal basis of  $GF(2^m)$  when  $m = p^n$  with prime  $p$ . Berlekamp (Ref. 12, p. 254), and Lidl and Niederreiter (Ref. 13, p. 124) have also

given a formula to compute the number of elements which can generate a normal basis in  $GF(2^m)$ . Wah and Wang (Refs. 14, 16) have shown that the so-called all-one-polynomial of degree  $m$  is irreducible and its roots constitute a normal basis if and only if  $m + 1$  is a prime and 2 is primitive mod  $(m + 1)$ . Pei, Wang and Omura (Ref. 15) have also presented necessary and sufficient conditions for an element to generate a normal basis in the field  $GF(2^m)$  for the case that  $m = 2^k p^n$  where  $p$  is an odd prime,  $k$  is a non-negative integer,  $n$  is a positive integer and  $p^n$  has 2 as one of its primitive roots. These conditions can be used to find a normal basis in  $GF(2^m)$ , if  $m$  is of the given form. Using this normal basis as the roots, one can construct an irreducible polynomial of degree  $m$  and, therefore, use the algorithm described in Ref. 10 to design the Massey-Omura multiplier.

In this article, a new algorithm to locate a normal basis in *any* field  $GF(2^m)$  is presented. In this algorithm, a special  $m \times m$  matrix needs to be set up and its nonsingularity needs to be verified. For large  $m$ , this algorithm also seems to be very time consuming. However, due to some special properties of this matrix, the matrix set up procedure only requires  $m$ , rather than  $m \times m$  entry computations, and, the verification of the nonsingularity can be based on some quick check rules, resulting in a saving of a tremendous amount of computation time. Using this normal basis, a methodology to construct the product function of the Massey-Omura multiplier is also developed in this article. This approach uses the concept of dual basis. It is shown that the coefficients of the product function are the trace values of some particular elements in  $GF(2^m)$ . These particular elements can be computed by the normal basis used and its dual basis. Hence, the design of a Massey-Omura multiplier can be based on any arbitrary normal basis in  $GF(2^m)$  which need not be the roots of the generator polynomial.

## II. Dual Basis Approach of Designing Massey-Omura Finite Field Multiplier

Two bases  $\{\alpha_1, \alpha_2, \dots, \alpha_m\}$  and  $\{\beta_1, \beta_2, \dots, \beta_m\}$  are said to be dual, or complementary, if

$$Tr(\alpha_i \beta_j) = \delta_{ij} = \begin{cases} 0 & \text{for } i \neq j \\ 1 & \text{for } i = j \end{cases} \quad (3)$$

Seven useful properties of a finite field  $GF(2^m)$  are stated here without proof (for proofs see Refs. 1 and 16). These properties are as follows:

- (1) Trace is a linear operation over  $GF(2)$ .

(2) For every  $\alpha \in GF(2^m)$

$$Tr(\alpha^2) = [Tr(\alpha)]^2 = Tr(\alpha) \in GF(2)$$

(3)  $Tr(1) = m \bmod 2$

(4) If  $\{\alpha_1, \alpha_2, \dots, \alpha_m\}$  and  $\{\beta_1, \beta_2, \dots, \beta_m\}$  are two bases and dual to each other, for any  $x \in GF(2^m)$ ,

$$x = \sum_{i=1}^m a_i \cdot \alpha_i$$

$$= \sum_{i=1}^m Tr(x \cdot \beta_i) \cdot \alpha_i$$

(5) Every basis has a dual basis.

(6) A normal basis exists in any field  $GF(2^m)$ .

(7) The dual basis of a normal basis is also a normal basis.

Suppose that  $\{\alpha, \alpha^2, \alpha^{2^2}, \dots, \alpha^{2^{m-1}}\}$  is a normal basis of field  $GF(2^m)$ , and  $\{\beta, \beta^2, \beta^{2^2}, \dots, \beta^{2^{m-1}}\}$  is its dual basis. For any two elements  $y$  and  $z$  in  $GF(2^m)$ , they can be expressed as

$$y = y_0 \alpha + y_1 \alpha^2 + y_2 \alpha^{2^2} + \dots + y_{m-1} \alpha^{2^{m-1}}$$

$$= \sum_{j=0}^{m-1} y_j \alpha^{2^j}$$

$$z = z_0 \alpha + z_1 \alpha^2 + z_2 \alpha^{2^2} + \dots + z_{m-1} \alpha^{2^{m-1}}$$

$$= \sum_{i=0}^{m-1} z_i \alpha^{2^i}$$

Let

$$\omega = y \circ z$$

$$= \omega_0 \alpha + \omega_1 \alpha^2 + \omega_2 \alpha^{2^2} + \dots + \omega_{m-1} \alpha^{2^{m-1}}$$

$$= \sum_{k=0}^{m-1} \omega_k \alpha^{2^k}$$

By property (4)

$$\begin{aligned} \omega_k &= Tr(\omega \beta^{2^k}) \\ &= Tr(yz \beta^{2^k}) \\ &= Tr\left(\sum_{i=0}^{m-1} y_i \alpha^{2^i} \sum_{j=0}^{m-1} z_j \alpha^{2^j} \cdot \beta^{2^k}\right) \\ &= \sum_{i=1}^{m-1} \sum_{j=0}^{m-1} y_i z_j Tr\left(\alpha^{2^i} \cdot \alpha^{2^j} \cdot \beta^{2^k}\right) \end{aligned} \quad (4)$$

**Lemma 1**

$$Tr\left(\alpha^{2^i} \cdot \alpha^{2^j} \cdot \beta^{2^k}\right) = Tr\left(\alpha^{2^{i-1}} \cdot \alpha^{2^{j-1}} \cdot \beta^{2^{k-1}}\right)$$

where

$$\bar{i} \triangleq i \bmod m$$

**Proof:** The lemma follows from the fact that  $x^{2^m} = x$  and  $Tr(x^2) = Tr(x)$  (Property 2) for any  $x \in GF(2^m)$ .

**Theorem 2**

$$\omega_{k-1} = \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} y'_i z'_j Tr\left(\alpha^{2^i} \cdot \alpha^{2^j} \cdot \beta^{2^k}\right)$$

where

$$y'_i = y_{\bar{i-1}}$$

**Proof:** From Eq. (4),

$$\begin{aligned} \omega_{k-1} &= \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} y_i z_j Tr\left(\alpha^{2^i} \cdot \alpha^{2^j} \cdot \beta^{2^{k-1}}\right) \\ &= \sum_{i=1}^m \sum_{j=1}^m y_{i-1} z_{j-1} Tr\left(\alpha^{2^{i-1}} \cdot \alpha^{2^{j-1}} \cdot \beta^{2^{k-1}}\right) \\ &= \sum_{i=1}^m \sum_{j=1}^m y'_{i-1} z'_{j-1} Tr\left(\alpha^{2^i} \cdot \alpha^{2^j} \cdot \beta^{2^k}\right) \end{aligned}$$

Since

$$y'_0 = y_{m-1} \text{ and } z'_0 = z_{m-1}$$

$$\omega_{k-1} = \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} y'_i z'_j \operatorname{Tr}(\alpha^{2^i} \cdot \alpha^{2^j} \cdot \beta^{2^k})$$

Let  $f$  be a  $2m$ -dimensional function such that

$$\begin{aligned} \omega_{m-1} &= f(y_0, y_1, \dots, y_{m-1}; z_0, z_1, \dots, z_{m-1}) \\ &\triangleq \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} y_i z_j \operatorname{Tr}(\alpha^{2^i} \cdot \alpha^{2^j} \cdot \beta^{2^{m-1}}) \\ &= \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} \rho_{ij} y_i z_j \end{aligned}$$

where  $\rho_{ij} \triangleq \operatorname{Tr}(\alpha^{2^i} \cdot \alpha^{2^j} \cdot \beta^{2^{m-1}})$ . Since  $\{y'_i\}$  is the cyclically shifted version of  $\{y_i\}$ , by Theorem 2,

$$\begin{aligned} \omega_{m-2} &= \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} y'_i z'_j \operatorname{Tr}(\alpha^{2^i} \cdot \alpha^{2^j} \cdot \beta^{2^{m-1}}) \\ &= \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} \rho_{ij} y'_i z'_j \\ &= f(y_{m-1}, y_0, y_1, \dots, y_{m-2}; \\ &\quad z_{m-1}, z_0, z_1, \dots, z_{m-2}) \end{aligned}$$

Applying this technique repeatedly, one can obtain

$$\left. \begin{aligned} \omega_{m-1} &= f(y_0, y_1, y_2, \dots, y_{m-1}; \\ &\quad z_0, z_1, z_2, \dots, z_{m-1}) \\ \omega_{m-2} &= f(y_{m-1}, y_0, y_1, \dots, y_{m-2}; \\ &\quad z_{m-1}, z_0, z_1, \dots, z_{m-2}) \\ &\vdots \\ \omega_1 &= f(y_2, y_3, \dots, y_{m-1}, y_0, y_1; \\ &\quad z_2, z_3, \dots, z_{m-1}, z_0, z_1) \\ \omega_0 &= f(y_1, y_2, \dots, y_{m-1}, y_0, \\ &\quad z_1, z_2, \dots, z_{m-1}, z_0) \end{aligned} \right\} \quad (5)$$

where

$$f(a_0, a_1, \dots, a_{m-1}; b_0, b_1, \dots, b_{m-1}) = \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} \rho_{ij} a_i b_j \quad (5a)$$

with

$$\rho_{ij} = \operatorname{Tr}(\alpha^{2^i} \cdot \alpha^{2^j} \cdot \beta^{2^{m-1}}) \quad (5b)$$

In Eq. (5), it is shown that the product components  $\omega_i$ ,  $i = 0, 1, \dots, m-1$ , can be obtained by the same logic function  $f$  operating on the cyclically shifted versions of the components of multiplicand and multiplier. This function  $f$ , the so-called product function, defines the Massey-Omura multiplier (Refs. 9, 10). It is illustrated from Eqs. (5), (5a) and (5b) that the product function  $f$  only depends on the normal basis  $\{\alpha, \alpha^2, \dots, \alpha^{2^{m-1}}\}$  used since  $\beta$  also depends on  $\alpha$  and that its coefficients are the trace values of some elements in  $GF(2^m)$ . These elements can be computed by multiplying the components of the normal basis and the last component  $\beta^{2^{m-1}}$  of its dual basis. Unlike the method described in Ref. 10, the construction of the product function  $f$  in this article is independent of the characteristic of the generating polynomial of  $GF(2^m)$ . Hence, this method gives an advantage that one can use any *arbitrary* irreducible polynomial of degree  $m$  to generate the field  $GF(2^m)$ .

### III. Properties of the Associated Boolean Matrix

An equivalent way to represent the product function  $f$  of Eq. (5) is by means of a Boolean matrix

$$\Omega \triangleq [\rho_{ij}]_{i,j=0}^{m-1} \quad (6)$$

where

$$\rho_{ij} = \operatorname{Tr}(\alpha^{2^i} \cdot \alpha^{2^j} \cdot \beta^{2^{m-1}}) \quad (6a)$$

is the coefficient of  $a_i b_j$  in Eq. (5a).

Since to design a Massey-Omura multiplier is essentially to design a product function, the construction of the Boolean matrix  $\Omega$  in Eq. (6) becomes the central issue of the design. The following theorems show some properties of the Boolean matrix.

**Theorem 3.** The matrix  $\Omega$  is symmetric, that is,  $\rho_{ij} = \rho_{ji}$ .

**Proof.** This is obvious from the definition of  $\rho_{ij}$  in Eq. (6a).

**Theorem 4**

$$\rho_{ii} = \begin{cases} 0, & \text{if } i \neq m-2 \\ 1, & \text{if } i = m-2 \end{cases}$$

**Proof**

$$\begin{aligned} \rho_{ii} &= Tr(\alpha^{2^i} \cdot \alpha^{2^i} \cdot \beta^{2^{m-1}}) \\ &= Tr(\alpha^{2^{i+1}} \cdot \beta^{2^{m-1}}) \\ &= \delta_{i(m-2)} \\ &= \begin{cases} 0, & \text{if } i \neq m-2 \\ 1, & \text{if } i = m-2 \end{cases} \end{aligned}$$

**Theorem 5**

$$\sum_{i=0}^{m-1} \rho_{ij} = \begin{cases} 0, & j \neq m-1 \\ 1, & j = m-1 \end{cases}$$

**Proof**

$$\begin{aligned} \sum_{i=0}^{m-1} \rho_{ij} &= Tr\left(\sum_{i=0}^{m-1} \alpha^{2^i} \cdot \alpha^{2^j} \cdot \beta^{2^{m-1}}\right) \\ &= Tr(\alpha^{2^j} \cdot \beta^{2^{m-1}}) \\ &= \delta_{j(m-1)} \\ &= \begin{cases} 0, & j \neq m-1; \\ 1, & j = m-1 \end{cases} \end{aligned}$$

From theorems 3 and 4, one can conclude that in the Boolean matrix  $\Omega$  only the  $(m^2 - m)/2$  entry values in the upper-right triangular portion must be computed, with the diagonal values being fixed. Theorem 5 shows that there are an odd number of 1's in the last row and last column, and, an even number of 1's in the remaining rows and columns. This gives a very simple check on the correctness of the Boolean matrix.

Once the Boolean matrix is formed, the design of the Massey-Omura multiplier can proceed as described in Ref. 10. However, there are two important issues which must be addressed.

(1) How can one find a normal basis in  $GF(2^m)$  if the generating irreducible polynomial of this field does not provide linearly independent roots?

(2) How can one find the dual basis of a normal basis in  $GF(2^m)$ ? In the next section we will address these two issues.

#### IV. Locating a Normal Basis in $GF(2^m)$

Suppose that  $\{\alpha, \alpha^2, \alpha^4, \dots, \alpha^{2^{m-1}}\}$  is a normal basis in  $GF(2^m)$ . From properties (5) and (7) of section II, its dual basis  $\{\beta, \beta^2, \dots, \beta^{2^{m-1}}\}$ , which is also normal, must exist in  $GF(2^m)$ . Since  $\{\beta, \beta^2, \dots, \beta^{2^{m-1}}\}$  is a basis of  $GF(2^m)$ , element  $\alpha$  can be expressed as

$$\alpha = a_0 \beta + a_1 \beta^2 + \dots + a_{m-1} \beta^{2^{m-1}} \quad (7)$$

where  $a_j \in GF(2)$  for  $j = 0, 1, \dots, m-1$ . By squaring Eq. (7) repeatedly and applying the property that  $\beta^{2^m} = \beta$ , one can write

$$\begin{bmatrix} \alpha \\ \alpha^2 \\ \alpha^4 \\ \vdots \\ \vdots \\ \alpha^{2^{m-1}} \end{bmatrix} = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_{m-1} \\ a_{m-1} & a_0 & a_1 & \cdots & a_{m-2} \\ a_{m-2} & a_{m-1} & a_0 & \cdots & a_{m-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & a_3 & \cdots & a_0 \end{bmatrix} \begin{bmatrix} \beta \\ \beta^2 \\ \beta^4 \\ \vdots \\ \vdots \\ \beta^{2^{m-1}} \end{bmatrix}$$

$$\triangleq A \cdot \begin{bmatrix} \beta \\ \beta^2 \\ \beta^4 \\ \vdots \\ \vdots \\ \beta^{2^{m-1}} \end{bmatrix} \quad (8)$$

Multiplying both sides of Eq. (8) by a row vector  $[\alpha \alpha^2 \alpha^4 \cdots \alpha^{2^{m-1}}]$ , we have

$$\begin{bmatrix} \alpha^2 & \alpha^3 & \alpha^5 & \cdots & \alpha^{(2^{m-1}+1)} \\ \alpha^3 & \alpha^4 & \alpha^6 & \cdots & \alpha^{(2^{m-1}+2)} \\ \alpha^5 & \alpha^6 & \alpha^8 & \cdots & \alpha^{(2^{m-1}+4)} \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ \alpha^{(2^{m-1}+1)} & \alpha^{(2^{m-1}+2)} & \alpha^{(2^{m-1}+4)} & \cdots & \alpha^{2^m} \end{bmatrix}$$

$$= \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_{m-1} \\ a_{m-1} & a_0 & a_1 & \cdots & a_{m-1} \\ a_{m-2} & a_{m-1} & a_0 & \cdots & a_{m-3} \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ a_1 & a_2 & a_3 & \cdots & a_0 \end{bmatrix}$$

$$\cdot \begin{bmatrix} \beta\alpha & \beta\alpha^2 & \beta\alpha^4 & \cdots & \beta\alpha^{2^{m-1}} \\ \beta^2\alpha & \beta^2\alpha^2 & \beta^2\alpha^4 & \cdots & \beta^2\alpha^{2^{m-1}} \\ \beta^4\alpha & \beta^4\alpha^2 & \beta^4\alpha^4 & \cdots & \beta^4\alpha^{2^{m-1}} \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ \beta^{2^{m-1}}\alpha & \beta^{2^{m-1}}\alpha^2 & \beta^{2^{m-1}}\alpha^4 & \cdots & \beta^{2^{m-1}}\alpha^{2^{m-1}} \end{bmatrix} \quad (9)$$

Since the trace function is linear over  $GF(2)$ , applying the trace function to both sides of Eq. (9) results in

$$\mathbf{F} = \mathbf{A} \cdot \mathbf{I} = \mathbf{A} \quad (10)$$

where  $\mathbf{F}$  is an  $m$  by  $m$  matrix with  $(i, j)$  entry

$$F_{ij} = Tr(\alpha^{2^i} \cdot \alpha^{2^j}) \quad (11)$$

for  $i, j = 0, 1, 2, \dots, m - 1$ . Notice that  $\mathbf{F}$  depends only on  $\alpha$ . Hence Eq. (8) can be written as

$$\begin{bmatrix} \alpha \\ \alpha^2 \\ \alpha^4 \\ \vdots \\ \vdots \\ \alpha^{2^{m-1}} \end{bmatrix} = \mathbf{F}(\alpha) \cdot \begin{bmatrix} \beta \\ \beta^2 \\ \beta^4 \\ \vdots \\ \vdots \\ \beta^{2^{m-1}} \end{bmatrix} \quad (12)$$

**Theorem 6.** For  $\alpha \in GF(2^m)$ ,  $\alpha, \alpha^2, \alpha^4, \dots, \alpha^{2^{m-1}}$  are linearly independent if and only if  $\mathbf{F}(\alpha)$  is invertible.

**Proof.** Only the proof of the sufficient condition is necessary since the proof of the necessary condition is trivial and well known. If  $\alpha, \alpha^2, \alpha^4, \dots, \alpha^{2^{m-1}}$  are linearly dependent, there exist  $c'_i$ 's,  $i = 0, 1, \dots, m - 1$  in  $GF(2)$  which are not all zeros such that

$$\sum_{i=0}^{m-1} c'_i \alpha^{2^i} = 0$$

Multiplying both sides by  $\alpha^{2^j}$  for  $j = 0, 1, \dots, m - 1$ ,

$$\sum_{i=0}^{m-1} c'_i \alpha^{2^i} \cdot \alpha^{2^j} = 0$$

Taking the trace values on both sides, one has

$$\sum_{i=0}^{m-1} c'_i Tr(\alpha^{2^i} \cdot \alpha^{2^j}) = \sum_{i=0}^{m-1} c'_i F_{ij} = 0$$

for all  $j = 0, 1, 2, \dots, m - 1$ . Thus,  $\mathbf{F}$  is not invertible and the theorem is proved.

From Eq. (12), if  $\mathbf{F}$  is invertible, the dual basis  $\{\beta, \beta^2, \dots, \beta^{2^{m-1}}\}$  of the normal basis  $\{\alpha, \alpha^2, \dots, \alpha^{2^{m-1}}\}$  can be computed by

$$\begin{bmatrix} \beta \\ \beta^2 \\ \vdots \\ \cdot \\ \beta^{2^{m-1}} \end{bmatrix} = \mathbf{F}^{-1}(\alpha) \cdot \begin{bmatrix} \alpha \\ \alpha^2 \\ \vdots \\ \cdot \\ \alpha^{2^{m-1}} \end{bmatrix} \quad (13)$$

resulting in a saving of a significant amount of computation time. The following theorems describe these conditions.

**Theorem 7.** If  $\text{Tr}(\alpha) = 0$ ,  $\mathbf{F}(\alpha)$  is not invertible.

**Proof.** This is obvious because  $\text{Tr}(\alpha) = 0$  implies that  $\{\alpha, \alpha^2, \alpha^4, \dots, \alpha^{2^{m-1}}\}$  are linearly dependent.

**Theorem 8.** If  $\text{Tr}(\alpha \alpha^{2^i}) = 1$  for all  $i = 0, 1, \dots, m-1$ ,  $\mathbf{F}(\alpha)$  is not invertible.

**Proof.** Since  $F_{\bar{i+1}, \bar{i+1}}(\alpha) = F_{ij}(\alpha)$ , the condition of  $\text{Tr}(\alpha \alpha^{2^i}) = 1$  for all  $i$  results in an all-one matrix  $\mathbf{F}(\alpha)$  which is not invertible.

**Theorem 9.** If  $m$  is even and  $\text{Tr}(\alpha) = \text{Tr}(\alpha^{2^{m/2}} + 1)$ ,  $\mathbf{F}(\alpha)$  is not invertible. The following two lemmas are required to prove this theorem.

**Lemma 10.** If the first row of matrix  $\mathbf{F}(\alpha)$  has an even number of 1's,  $\mathbf{F}(\alpha)$  is not invertible.

**Proof.** Since the  $(i+1)$ th row of  $\mathbf{F}(\alpha)$  is the cyclically shifted version of the  $i$ th row, this condition means that  $\mathbf{F}(\alpha)$  has an even number of 1's in all rows. Adding up all column vectors results in an all-zero vector. Hence  $\mathbf{F}(\alpha)$  is not invertible.

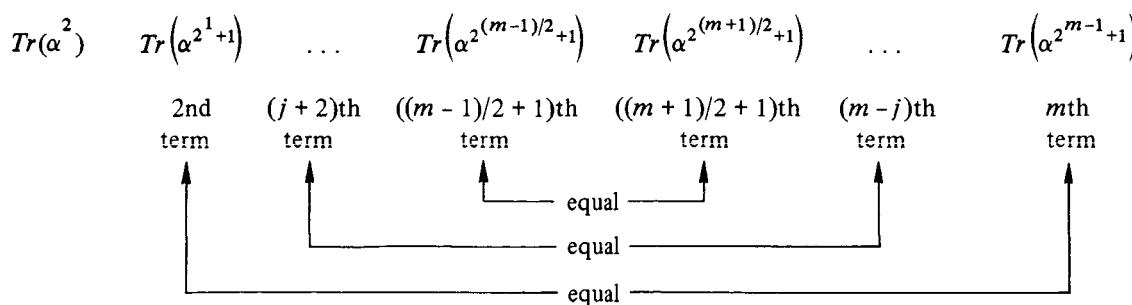
**Lemma 11.**  $\text{Tr}(\alpha^{2^j+1}) = \text{Tr}(\alpha^{2^{m-j}+1})$  for  $1 \leq j < m/2$ .

**Proof**

$$\begin{aligned} \text{Tr}(\alpha^{2^{m-j}+1}) &= \text{Tr}\left[\left(\alpha^{2^{m-j}+1}\right) z^j\right] = \text{Tr}\left(\alpha^{2^m+2^j}\right) \\ &= \text{Tr}\left(\alpha^{2^j+1}\right) \end{aligned}$$

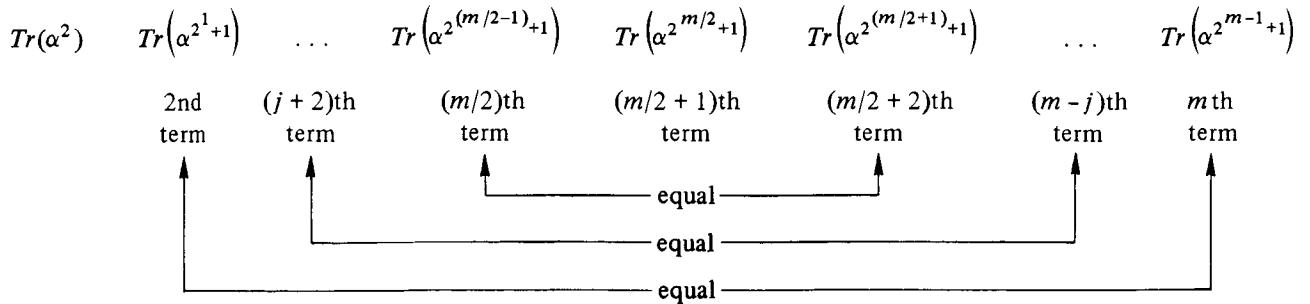
Lemma 11 implies that the  $(j+1)$ th element from the left of the first row vector of  $\mathbf{F}(\alpha)$  is equal to the  $j$ th element from the right. Lemma 10 and Lemma 11 lead to the following two properties.

(1) When  $m$  is odd, the first row vector of the matrix  $\mathbf{F}(\alpha)$  has the structure



Therefore if the first element  $\text{Tr}(\alpha^2) = 0$ ,  $\mathbf{F}(\alpha)$  is not invertible. This is equivalent to the Theorem 7 since  $\text{Tr}(\alpha) = \text{Tr}(\alpha^2)$ .

(2) When  $m$  is even, the structure of the first row of  $\mathbf{F}(\alpha)$  becomes



This implies that if  $\text{Tr}(\alpha) = \text{Tr}(\alpha^{2^{m/2+1}})$ ,  $\mathbf{F}(\alpha)$  is not invertible since there are an even number of 1's in the first row. Hence, Theorem 9 is proved.

## VI. Example and Results

This section gives an example of designing the Massey-Omura multiplier for  $GF(2^7)$ . Let  $P(x) = x^7 + x^3 + 1$  be the generating polynomial of  $GF(2^7)$ . Suppose that  $\theta$  is a root of  $P(x)$ . Note that the roots  $\{\theta, \theta^2, \theta^4, \theta^{2^3}, \theta^{2^4}, \theta^{2^5}, \theta^{2^6}\}$  are not linearly independent. By Eq. (A-3), the trace values of the canonical basis,  $\{\text{Tr}(\theta^i) | i = 0, 1, \dots, 6\}$  are  $\{1, 0, 0, 0, 0, 0, 0\}$ .

- (1) Let  $\alpha = \theta$ . By Theorem 7,  $\mathbf{F}(\alpha)$  is not invertible since  $\text{Tr}(\alpha) = 0$ .
- (2) Let  $\alpha = 1 + \theta$ . Then  $\text{Tr}(\alpha \cdot \alpha^{2^i}) = 1$  for all  $i = 0, 1, \dots, m-1$ . By Theorem 8,  $\mathbf{F}(\alpha)$  is not invertible.
- (3) Let  $\alpha = \theta^2$ .  $\text{Tr}(\alpha) = 0$  and  $\mathbf{F}(\alpha)$  is not invertible.
- (4) Let  $\alpha = 1 + \theta^2$ .  $\text{Tr}(\alpha \cdot \alpha^{2^i}) = 1$  for all  $i = 0, \dots, m-1$ . The matrix  $\mathbf{F}(\alpha)$  is not invertible.
- (5) Let  $\alpha = \theta + \theta^2$ .  $\text{Tr}(\alpha) = 0$ . The matrix  $\mathbf{F}(\alpha)$  is not invertible.
- (6) Let  $\alpha = 1 + \theta + \theta^2$ .  $\text{Tr}(\alpha \cdot \alpha^{2^i}) = 1$  for all  $i = 0, \dots, m-1$ . The matrix  $\mathbf{F}(\alpha)$  is not invertible.
- (7) Let  $\alpha = \theta^3$ .  $\text{Tr}(\alpha) = 0$ . The matrix  $\mathbf{F}(\alpha)$  is not invertible.
- (8) Let  $\alpha = 1 + \theta^3$ .

$$\mathbf{F}(\alpha) = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

is invertible and its inverse is given by

$$\mathbf{F}^{-1}(\alpha) = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix}$$

Hence  $\{\alpha, \alpha^2, \dots, \alpha^{2^{m-1}}\}$  is a normal basis

- (9) The dual basis  $\{\beta\}$  of  $\{\alpha\}$  is given by

$$\begin{bmatrix} \beta \\ \beta^2 \\ \beta^4 \\ \vdots \\ \beta^{2^6} \end{bmatrix} = \mathbf{F}^{-1}(\alpha) \cdot \begin{bmatrix} \alpha \\ \alpha^2 \\ \alpha^4 \\ \vdots \\ \alpha^{2^6} \end{bmatrix}$$

Therefore

$$\begin{aligned} \beta &= \alpha + \alpha^2 + \alpha^8 + \alpha^{16} + \alpha^{64} \\ &= 1 + \theta + \theta^3 + \theta^6 \end{aligned}$$

(10) Finally,

$$\Omega = [Tr(\alpha^{2i} \cdot \alpha^{2j} \cdot \beta^{2m-1})]_{i,j=0}^{m-1}$$

$$= \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

Once the Boolean matrix is constructed, the product function is defined. Then the implementation of the Massey-Omura multiplier of  $GF(2^7)$  can be designed as described in Ref. 10. Figures 2–4 give Boolean matrices for  $m = 8, 17$  and  $30$ , respectively. Reference 16 also gives a Boolean matrix for  $m = 127$ . It should be pointed out that, in our experience of searching the Boolean matrix, the above-mentioned three quick ways as to verifying the invertibility of  $\mathbf{F}(\alpha)$  given in Theorems 7 through 9 are the primary verification rules that the procedure has gone through. In other words, our experience indicates that, in the process of constructing the Boolean matrix shown in Fig. 1, the most time-consuming matrix inversion procedure in the Gaussian elimination method is unlikely to be needed to rule out the candidate  $\alpha$ , resulting in

a saving of a great deal of time. Figure 5 illustrates the CPU time required to construct the Boolean matrix for  $GF(2^m)$  on a VAX-11/750. The capital delta ( $\Delta$ ) in the figure indicates the actual time required by using an arbitrarily selected irreducible polynomial of degree  $m$ . For example, the construction of the Boolean matrix for  $GF(2^{127})$  takes only 40 minutes. The solid line shows that the trend of the required time increases exponentially as  $m$  increases. For large  $m$ , the computation time is mainly for forming the Boolean matrix, while, for small  $m$ , the computation time is dominated by the pre-matrix computation including the initial program set up and the trace computations of canonical basis which is required for forming matrix  $\mathbf{F}(\alpha)$ . The most vertical part of the line in Fig. 5 shows the transition between these two kinds of computation.

## VII. Conclusion

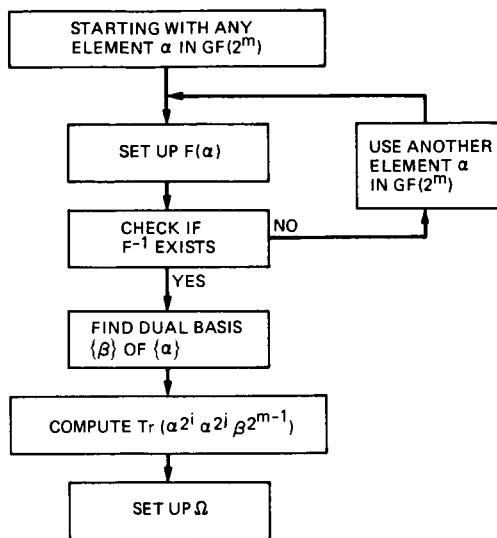
Although for some Galois field  $GF(2^m)$  the roots of a generating polynomial can be easily verified to be linearly independent and then used as a normal basis, it is generally very difficult to locate a normal basis in a field. This makes the Massey-Omura multiplication less attractive since its design is based on a normal basis. A generalized algorithm to locate a normal basis of  $GF(2^m)$  has been presented. Using this normal basis, an algorithm to construct a product function has also been developed. After a product function is defined, the design of the Massey-Omura multiplier is straightforward.

## Acknowledgment

The author thanks Mr. D. Y. Pei for his contribution in helping with the Appendix.

## References

1. MacWilliams, F. J., and Sloane, N. J. A., *The Theory of Error-Correcting Codes*, North-Holland Publishing, New York, 1977.
2. Peterson, W. W., and Weldon, Jr., E. J., *Error-Correcting Codes*, MIT Press, Cambridge, 1972.
3. Berkovits, S., Kowalchuk, J., and Schanning, B., "Implementing Public Key Scheme," *IEEE Communications Magazine* 17, pp. 2-3, May 1979.
4. Yeh, C. S., Reed, I. S., and Troung, T. K., "Systolic Multipliers for Finite Fields  $GF(2^m)$ ," *IEEE Transactions on Computers C-33*, No. 4, pp. 357-360, April 1984.
5. Bartee, T. C., and Schneider, D. I., "Computation with Finite Fields," *Inform. Contr.* 6, pp. 79-98, Mar. 1963.
6. Gallaghert, R. G., *Information Theory and Reliable Communication*, New York: Wiley, 1968.
7. Laws, B. A., and Rushforth, C. K., "A Cellular-array Multiplier for  $GF(2^m)$ ," *IEEE Transactions on Computers C-20*, pp. 1573-1578, Dec. 1971.
8. Kung, H. T., "Why Systolic Architectures?" *IEEE Computer* 15, pp. 37-46, Jan. 1982.
9. Massey, J. L., and Omura, J. K., U. S. Patent Application of "Computational Method and Apparatus for Finite Field Arithmetic," submitted in 1981.
10. Wang, C. C., et al., "VLSI Architectures for Computing Multiplications and Inverses in  $GF(2^m)$ ," *IEEE Transactions on Computers C-34*, No. 8, pp. 709-717, August 1985.
11. Perlis, S., "Normal Basis of Cyclic Fields of Prime-Power Degree," *Duke Math. J.* 9, pp. 507-517, 1942.
12. Berlekamp, E. R., *Algebraic Coding Theory*, McGraw-Hill Book Company, 1968.
13. Lidl, R., and Niederreiter, H., *Finite Fields*, Addison-Wesley Publishing Company, 1983.
14. Wah, P. K. S., and Wang, M. Z., "Realization and Application of the Massey-Omura Lock," *Proceedings of International Zurich Seminar*, IEEE press, pp. 175-182, March 1984.
15. Pei, D. Y., Wang, C. C., and Omura, J. K., "Normal Basis of Finite Field  $GF(2^m)$ ," *IEEE Transactions on Information Theory IT-32*, No. 2, pp. 285-287, March 1986.
16. Wang, C. C., "Exponentiation in Finite Field  $GF(2^m)$ ," Ph.D. dissertation, School of Engineering and Applied Sciences, UCLA, February 1985.



**Fig. 1. Algorithm of constructing the Boolean matrix for the multiplication in  $GF(2^m)$**

:	0	1	2	3	4	5	6	7
0 :	0	1	0	0	1	1	1	0
1 :	1	0	1	1	0	0	0	1
2 :	0	1	0	1	1	1	0	0
3 :	0	1	1	0	0	0	0	0
4 :	1	0	1	0	0	1	0	1
5 :	1	0	1	0	1	0	0	1
6 :	1	0	0	0	0	0	1	0
7 :	0	1	0	0	1	1	0	0

NUMBER OF 1 IN BOOLEAN MATRIX = 27

**Fig. 2. Boolean matrix for  $GF(2^8)$**

:	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
0 :	0	1	1	1	0	0	1	1	1	0	0	1	1	1	1	0
1 :	1	0	1	1	0	0	0	0	1	0	0	1	0	1	1	0
2 :	1	1	0	1	0	0	0	1	0	0	1	1	1	1	1	0
3 :	1	1	1	0	0	0	0	1	0	0	1	1	1	0	1	0
4 :	0	0	0	0	0	0	0	1	1	0	1	0	1	1	1	0
5 :	0	0	0	0	0	0	0	1	1	1	1	0	0	1	1	1
6 :	1	0	0	0	0	0	0	1	1	1	0	0	0	1	0	1
7 :	1	0	1	1	1	1	1	0	1	1	0	0	0	1	1	1
8 :	1	1	0	0	1	1	1	0	1	0	0	0	1	0	1	1
9 :	0	0	0	0	0	1	1	1	0	1	0	0	0	1	0	0
10 :	0	0	1	1	1	1	0	0	0	1	0	1	1	0	0	1
11 :	1	1	1	1	0	0	0	0	0	1	0	1	1	0	0	1
12 :	1	0	1	1	1	0	0	0	0	0	1	1	0	0	0	0
13 :	1	1	1	0	1	1	1	1	0	0	1	0	0	0	1	0
14 :	1	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0
15 :	0	0	1	0	0	1	0	1	1	0	0	0	0	1	0	1
16 :	0	1	0	0	0	1	1	1	0	1	1	0	0	0	0	0

NUMBER OF 1 IN BOOLEAN MATRIX = 137

**Fig. 3. Boolean matrix for  $GF(2^{17})$**

:	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2				
:	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
0 :	0	1	0	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	1	0	0	1	0	0	1	0	0	1		
1 :	1	0	0	0	0	1	0	0	1	0	0	1	0	1	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	
2 :	0	0	0	0	0	0	0	1	0	1	1	1	0	0	0	1	1	0	0	1	1	0	1	0	1	1	0			
3 :	1	0	0	0	1	0	1	0	1	1	1	1	1	1	1	0	0	1	1	0	1	0	0	1	0	1	0			
4 :	1	0	0	1	0	0	1	1	0	1	1	0	0	0	1	0	0	1	1	0	1	0	1	0	0	1	0			
5 :	1	1	0	0	0	0	1	0	0	0	1	1	0	1	1	1	0	0	1	1	0	0	1	0	0	0	0			
6 :	1	0	0	1	1	0	0	1	1	0	1	1	0	0	1	0	0	1	1	0	1	0	0	1	1	0	0			
7 :	0	0	0	0	1	1	1	0	0	0	1	0	0	1	1	0	0	0	1	0	1	1	0	0	0	0	0			
8 :	1	1	1	1	0	0	1	0	0	1	1	0	0	0	1	1	0	0	0	1	0	1	0	0	0	1	0			
9 :	1	0	0	1	1	0	0	0	1	0	1	1	1	1	0	1	1	1	0	0	0	1	0	1	0	1	0			
10 :	1	0	1	1	1	0	1	1	1	1	0	0	0	1	1	0	0	1	1	0	0	0	1	1	1	1	0			
11 :	1	1	1	1	0	1	1	0	0	0	0	1	0	0	1	1	1	0	1	0	0	0	1	0	0	0	1			
12 :	0	0	1	1	0	1	1	1	0	1	0	0	1	0	1	1	0	1	1	1	1	0	0	0	0	1	0			
13 :	1	0	1	1	0	0	0	0	1	1	0	0	0	1	1	1	0	1	1	1	0	1	0	1	1	0	0			
14 :	1	1	0	1	1	1	0	0	1	1	0	1	1	0	0	0	0	1	1	1	1	0	1	1	1	1	0			
15 :	1	0	0	1	0	0	1	1	1	0	1	0	1	0	0	0	1	0	0	0	1	1	0	1	1	0				
16 :	0	1	0	1	0	1	1	1	0	1	1	1	0	0	0	1	0	0	0	1	1	1	0	0	1	0	1			
17 :	1	1	0	1	1	1	0	0	0	0	1	1	1	0	0	1	1	0	1	1	0	1	0	1	1	0	1			
18 :	1	1	1	0	1	1	1	0	0	1	1	1	0	1	0	0	0	1	0	1	1	1	0	1	1	1	1			
19 :	0	0	1	0	0	0	1	0	1	1	0	0	1	1	1	0	1	0	0	0	0	1	0	1	1	1	1			
20 :	1	0	0	1	1	0	1	0	1	1	1	0	0	1	0	0	0	0	1	0	1	0	1	0	1	0	1			
21 :	1	0	0	1	1	1	0	0	1	0	0	1	1	1	0	1	0	1	0	0	0	0	1	0	1	0	0			
22 :	1	0	1	0	0	1	1	1	0	0	1	0	1	0	1	0	1	0	0	0	0	1	0	1	1	0				
23 :	1	0	1	1	1	0	0	1	1	0	0	0	0	1	0	0	1	0	0	0	0	1	1	0	1	0	0			
24 :	0	1	0	0	0	0	0	0	0	1	1	0	0	1	1	0	1	1	1	1	0	1	0	0	1	1	0			
25 :	0	0	1	0	1	1	1	0	1	1	0	0	1	1	0	0	1	0	1	0	1	1	0	0	0	1	1			
26 :	1	0	0	1	0	0	1	1	0	0	1	0	0	1	1	0	1	1	0	1	0	0	1	0	0	0	0			
27 :	0	0	1	0	0	0	1	0	0	1	1	0	1	0	1	1	1	1	0	1	1	1	0	0	0	1				
28 :	0	0	1	1	1	0	0	0	1	0	0	0	1	1	0	1	1	0	0	1	0	0	1	0	0	1				
29 :	1	1	0	0	0	0	0	0	0	1	1	0	1	1	1	1	1	1	1	0	0	1	0	0	1	0	0			

NUMBER OF 1 IN BOOLEAN MATRIX = 443

Fig. 4. Boolean matrix for  $GF(2^{30})$

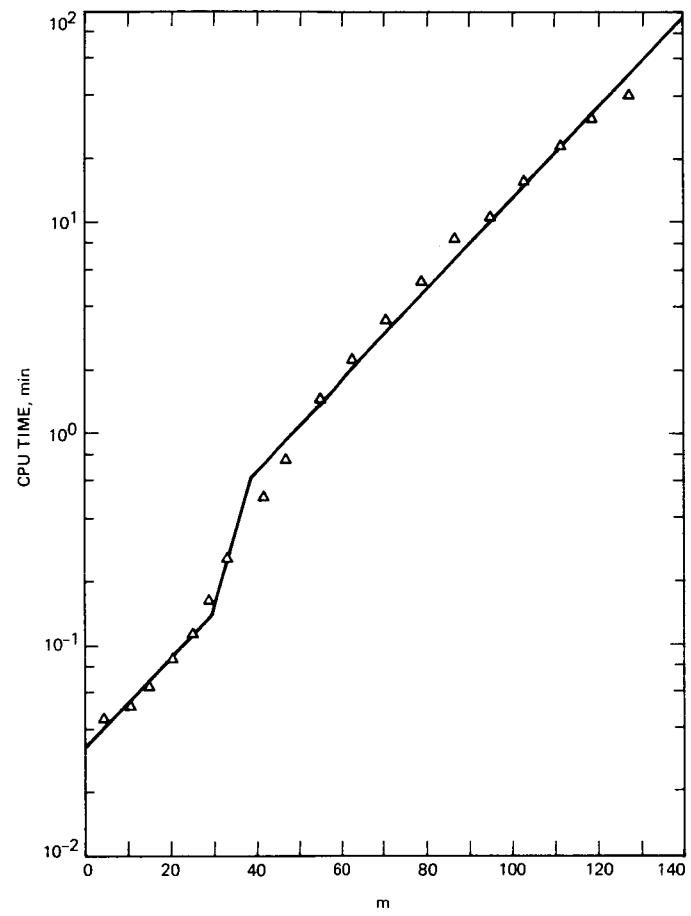


Fig. 5. CPU time required to construct Boolean matrix for  $GF(2^m)$

## Appendix A

### Trace Computation for $GF(2^m)$

Although the trace value of an element  $\theta$  in  $GF(2^m)$  can be computed directly by its definition

$$Tr(\theta) = \sum_{i=0}^{m-1} \theta^{2^i}$$

this appendix provides a much simpler approach to compute the trace value.

Let  $\epsilon$  be a root of the generating irreducible polynomial  $P(x)$  of  $GF(2^m)$ . For any element  $\theta$  of  $GF(2^m)$ ,

$$\begin{aligned} \theta &= a_0 + a_1 \epsilon + a_2 \epsilon^2 + \cdots + a_{m-1} \epsilon^{m-1} \\ &= \sum_{i=0}^{m-1} a_i \epsilon^i \end{aligned} \quad (\text{A-1})$$

Since trace is a linear operator in  $GF(2^m)$ ,

$$Tr(\theta) = \sum_{i=0}^{m-1} a_i Tr(\epsilon^i) \quad (\text{A-2})$$

Hence, our problem becomes how to find the trace values of the canonical basis  $\{1, \epsilon, \epsilon^2, \dots, \epsilon^{m-1}\}$  of  $GF(2^m)$ . The set of Newton formulae (Ref. A-1) demonstrates a very easy and quick way to accomplish it.

Let the generating polynomial be

$$P(X) = X^m + c_1 X^{m-1} + c_2 X^{m-2} + \cdots + c_{m-1} X + c_m$$

and  $\{\epsilon, \epsilon^2, \epsilon^4, \dots, \epsilon^{2^{m-1}}\}$  be the set of its roots. By Newton formulae, it can be shown that

$$\left. \begin{aligned} Tr(1) &= m \bmod 2 \\ Tr(\epsilon) + c_1 &= 0 \\ Tr(\epsilon^2) + c_1 Tr(\epsilon) &= 0 \\ &\vdots \\ Tr(\epsilon^j) + c_1 Tr(\epsilon^{j-1}) + \cdots & \\ + c_{j-1} Tr(\epsilon) + [j \bmod 2] c_j &= 0 \\ &\vdots \\ Tr(\epsilon^{m-1}) + c_1 Tr(\epsilon^{m-2}) + \cdots + c_{m-2} Tr(\epsilon) & \\ + [(m-1) \bmod 2] c_{m-1} &= 0 \end{aligned} \right\} \quad (\text{A-3})$$

Therefore, the trace values of the canonical basis  $\{1, \epsilon, \epsilon^2, \dots, \epsilon^{m-1}\}$  can be easily computed.

An interesting case is that when a trinomial is used to generate a field  $GF(2^m)$ . References A-2, A-3, and A-4 give a list of trinomials which are irreducible for  $m < 1000$ . In this case, the way of computing the trace values of the canonical basis  $\{1, \epsilon, \epsilon^2, \dots, \epsilon^{m-1}\}$  can be further simplified from Eq. (A-3) to the following:

Suppose that  $P(X) = X^m + X^k + 1$ . Let  $j \triangleq m - k$ .

(1)  $Tr(1) = m \bmod 2$ .

(2) When  $j$  is even,  $Tr(\epsilon^i) = 0$  for  $0 < i \leq m-1$ .

(3) When  $j$  is odd, for  $0 < i \leq m-1$ ,

$$Tr(\epsilon^i) = \begin{cases} 1, & \text{if } i = nj \text{ (} n \text{ is an integer)} \\ 0, & \text{otherwise} \end{cases} \quad (\text{A-4})$$

Note that if  $P(X)$  is irreducible and  $m$  is even,  $k$  and  $j$  must be odd.

## References

- A-1. Rédei, L., *Algebra*, Volume I, Pergamon Press, London, 1967.
- A-2. Zierler, N., and Brillhart, J., "On Primitive Trinomials (Mod 2)," *Information and Control* 13, pp. 541-554, 1968.
- A-3. Zierler, N., and Brillhart, J., "On Primitive Trinomials (Mod 2), II," *Information and Control* 14, pp. 566-569, 1969.
- A-4. Zierler, N., "On  $X^n + X + 1$  over  $GF(2)$ ," *Information and Control* 16, pp. 502-505, 1970.

D15-32  
Q-8

N87-15344

## X-Band Uplink Feedcone Capabilities, Components, and Layout

H. Marlin, A. Freiley, and R. Hartop  
Radio Frequency and Microwave Subsystems Section

*Two new X- (7.2 GHz up, 8.4 GHz down) and S-band (2.1 to 2.3 GHz) common aperture (XSC) feedcones are being added to the DSS 45 and DSS 65 34-Meter High Efficiency Antennas. These new feedcones are modifications of the existing SXC feedcone design incorporating a new high power (20-kW) X-band transmitter. The modified Antenna Microwave Subsystem design also incorporates two additional X-band low noise amplifiers and greater phase stability performance to meet both the increased stability requirements for Galileo gravity wave experiments and requirements for spacecraft navigation near the sun. A third XSC will be constructed for DSS 15 later.*

### I. Introduction

This article is the first of several describing the implementation and performance of the new X-band uplink (7.2 GHz) feedcones (XSC), two of which are now being assembled at JPL. One will be mounted on the new 34-m high-efficiency antenna at DSS 65, and the other will replace the SXC feedcone presently mounted on the DSS 45 antenna. The two existing SXC feedcones, designed and manufactured by Ford Aerospace, Palo Alto, CA, are repackaged versions of the first generation X/S listen-only feedcone (Ref. 1). Principal characteristics of the new XSC feedcone are X-band (7.2 GHz) 20-kW transmission, X-band diplexed operation, X-band (8.4 GHz down) RCP and LCP simultaneous reception, and subsystem monitor and control from the Local Monitor and Control (LMC) Center through the DSN Local Area Network (LAN). Figure 1 is a functional block diagram of the XSC feedcone. Blocks identified with a shaded corner square are

new/improved components or assemblies. The XSC feedcone was designed to meet the new, stringent X-band phase stability specifications for the Galileo mission.

A prototype X-band receive/transmit feedcone, a second generation X/S feedcone, has been operated at DSS 13 over the last few years (Refs. 1 and 2). Significant overall system stability testing has been performed on this system, and new system stability measurement equipment and techniques have been developed in conjunction with the testing (Ref. 3).

The new XSC configuration allows either X-band maser to be switched to diplexed or low-noise operation with either LCP or RCP polarization. A new Bethe-hole directional coupler located just before the feedhorn X-band input section allows injection of phase calibration signals to both X-band receive

polarizations at the highest possible point in the system. Table 1 gives a comparison of TDA-specified antenna microwave subsystem performance requirements with the current design. Table 2 provides system noise temperature estimates of the XSC configuration.

The DSS 65 feedcone is scheduled to arrive in Spain and be mounted on the antenna during the month of May 1987, with the antenna operational by January 1988. The DSS 45 antenna, with the new feedcone installed, will be returned to operations in February 1988.

## II. Layout

The existing SXC listen-only feedcones were designed with the anticipation of future addition of S- and X-band 20-kW dplexed transmit capability. The initial XSC layout effort consisted of adding two additional X-band low noise amplifiers, a new cooled S-band FET low noise amplifier, a new 20-kW X-band transmitter assembly, and new waveguide cooling manifolds to the existing SXC design. The initial XSC layout plan was based on the Ford SXC design, which was in turn a development of the existing DSS 13 feedcone. This design presented few serious access or maintenance problems. As the basic design was modified to meet the new X-band uplink stability requirements, much of the Frequency and Timing Subsystem (FTS) equipment and Receiver-Exciter Subsystem (RCV) equipment normally located elsewhere on the antenna now would apparently need to be located within the feedcone. The principal driver was the need to reduce temperature-induced phase changes in coaxial cables by reducing their length to a minimum. However, as the design effort progressed, it became increasingly apparent that the feedcone was becoming too crowded for proper or safe maintenance. A partial mock-up was constructed to allow operations as well as engineering personnel to evaluate the equipment layout in terms of maintenance. The design was determined to be unacceptable from a maintenance standpoint.

The subsystem requirements were reviewed, and a number of possible system configurations were devised that moved some of the equipment into the elevation room. After careful review, it was decided to move the Exciter (RCV) and X5 Frequency Multiplier/100-MHz Distribution (FTS) into the elevation room, leaving the transmitter in the feedcone. The phase-critical cables running between the elevation room and the feedcone would be placed in an air-conditioned duct. The resulting redesigned feedcone is now acceptable in terms of maintenance and operations. The addition of 20-kW S-band dplexed capability to the XSC feedcone should S-band transmission become a future requirement will be difficult in that the space allowed in the SXC feedcone design for the S-band diplexer and associated waveguide has of necessity been eliminated.

## III. Feedcone Components

### A. Broadband Polarizer

The quarterwave plate polarizer used in the SXC feedcones has insufficient bandwidth in terms of VSWR and axial ratio to operate over the needed 7145- to 8600-MHz range. A water-cooled iris type polarizer of the design tested in the second generation X/S feedcone (Ref. 4) is being fabricated for use in the XSC feedcone using the electroformed copper process.

### B. X-Band Diplexer

The diplexer, a modified version of the three-section diplexer used in the second generation X/S feedcone installed at DSS 13 (Ref. 4) has lower insertion loss, improved match and lower fabrication cost and is a single copper electroformed, T-shaped, three-port, water-cooled assembly. The diplexer is made up of three major circuits — the combiner section, a transmitter filter, and a receiver filter. The transmitter filter consists of a seven-cavity bandpass filter and a four-cavity bandstop filter designed to pass transmitter frequencies (7145 to 7235 MHz) and reject receiver frequencies (8200 to 8600 MHz). The receiver filter consists of a one-cavity bandpass filter, a reduced width waveguide high-pass filter, and a two-cavity bandstop filter; the filter is designed to pass receiver frequencies (8200 to 8600 MHz) and reject transmitter frequencies (7145 to 7235 MHz). Two production units have been fabricated. The diplexer isolation/rejection measurements are 102 dB at 7145 to 7235 MHz and 80 dB at 8200 to 8600 MHz.

### C. X-Band Preamplifier Filter

The X-Band Preamplifier Filter (XPF) (Ref. 5) is a five-cavity band reject filter (7145 to 7235 MHz) that is placed between the orthomode and the maser in the low-noise path to provide sufficient maser/X-band transmitter isolation as well as to protect the maser from X-band transmit frequency signals from external sources. The filter is fabricated from standard copper waveguide and is uncooled. The XPF measured rejection is >75 dB; measured insertion loss is <0.03 dB.

### D. X-Band Orthomode Junction

The orthomode junction is an improved water-cooled, high-power, broadband version of the SXC orthomode junction. It is fabricated using the copper electroform method. Testing of production orthomodes is now in progress.

### E. X-Band PCG Coupler

The PCG coupler, located between the feedhorn and the polarizer, provides a means of injecting phase calibration signals into the feed system (7900 to 8900 MHz). The coupler is a unique form of a Bethe-hole directional coupler using two

properly sized and excited opposed coupling holes in circular waveguide. It is necessary to maintain bilateral symmetry because the circular waveguide at this point in the feed system can support the propagation of higher order transverse magnetic (TM) modes. These TM modes could seriously degrade the illumination efficiency of the feed. A high-pass filter consisting of a section of reduced-width waveguide is used to protect the comb generator assembly (Frequency and Timing Subsystem) from coupled transmitter power. The fabrication of the first production unit is complete. Measured coupling is  $38.2 \pm 0.5$  dB. When installed, the PCG directional coupler is rotated 45 degrees relative to the principal planes of the orthomode, allowing the same signals to be injected into both orthogonal arms of the orthomode. The effective installed coupling value is therefore increased by 3 dB to  $41.2 \pm 0.5$  dB.

#### F. WR430 X-Band Reject Filter

The amount of leakage of X-band transmitter power into the S-band low noise amplifier through the combiner is uncertain. The X-band reject filter will increase the X- to S-band isolation by 60 dB. During testing of the feedcone, measure-

ments will be performed to determine if the filter can perhaps be removed and replaced with a straight section of waveguide. The filter is fabricated from standard WR430 waveguide and is designed for receive applications only. Fabrication has started, with delivery scheduled for 26 September 1986.

### IV. Status

The assembly of the feedcones is in progress in the High Bay Building 280. Feedhorns have been installed in both and the installation of the waveguide into cone SN 3 (for DSS 65) and SN 4 (DSS 45) has begun.

The addition of X-band uplink capability to DSS 15 is planned for late 1988. The SXC feedcone now mounted on DSS 45 will be returned to JPL after it is replaced by the XSC feedcone. It will be refurbished and upgraded to XSC capabilities, and then mounted on the DSS 15 antenna. The SXC feedcone now at DSS 15 will then be stored for future modifications as required.

## Acknowledgments

The authors would like to acknowledge Dr. Farzin Manshadi and Phil Stanton for microwave component design work, Gary Bury and Bob Verish for cone layout and subsystem interface design work, H. Reilly and Bill Folwell for test support, and Dan Vienna for procurement support.

## References

1. Williams, W., and Reilly, H., "A Prototype DSN X/S-Band Feed: DSS 13 Application Status (Fourth Report)," *TDA Progress Report 42-60*, December 15, 1980, Jet Propulsion Laboratory, Pasadena, Calif., pp. 77-88.
2. Hartop, R., Johns, C., and Kolby, R., "X-Band Uplink Ground Systems Development," *TDA Progress Report 42-56*, January and February, 1980, Jet Propulsion Laboratory, Pasadena, Calif., pp. 48-58.
3. Otoshi, T., "DSN End-to-End Stability Testing," *TDA Progress Report 42-86*, Jet Propulsion Laboratory, Pasadena, Calif., August 15, 1986, pp. 1-14.
4. Withington, J. R., "Second-Generation X/S Feedcone: Capabilities, Layout and Components," *TDA Progress Report 42-63*, March and April, 1981, Jet Propulsion Laboratory, Pasadena, Calif., pp. 97-103.
5. Manshadi, F., "X-Band Preamplifier Filter," *TDA Progress Report 42-86*, Jet Propulsion Laboratory, Pasadena, Calif., August 15, 1986, pp. 227-230.

**Table 1. Antenna Microwave Subsystem Performance Requirements vs Design**

Parameter	Requirement	Design
<b>Receiving:</b>		
S-band:		
Frequency range, MHz	2200 to 2300	2200 to 2300
Antenna gain, dBi	55.8 + 0.0, -0.5	55.8 + 0.0, -0.5 <sup>a</sup>
System noise temp:		
FET, K	<150	55 ± 10 <sup>b</sup>
Polarization	RCP/LCP remotely selectable	RCP/LCP remotely selectable
Ellipticity, dB p-p	1.0	1.0
X-band:		
Frequency range:		
Maser (Blk III) MHz	8400 to 8500	8400 to 8500
FET, MHz	8200 to 8600	8200 to 8600
Antenna gain, dBi	67.3 + 0.0, -0.8	67.3 + 0.0, -0.8 <sup>a</sup>
System noise temp:		
Maser (Blk II):		
Low noise, K	18.5 ± 2	TWM I: 20.5 TWM II: 24.5 <sup>c</sup>
Diplexed, K	21.5 ± 3	TWM I: 29.9 TWM II: 33.8 <sup>c</sup>
FET, K	<150	60 ± 10 <sup>d</sup>
Polarization	RCP/LCP remotely selectable	RCP/LCP remotely selectable
Ellipticity, dB p-p	0.7	0.7
<b>Transmitting:</b>		
S-band	Not required	None
X-band:		
Frequency range, MHz	7145 to 7335	7145 to 7335
Antenna gain, dBi	65.9 + 0.0, -0.8	65.9 + 0.0, -0.8 <sup>a</sup>
Polarization	RCP/LCP remotely selectable	RCP/LCP remotely selectable
Ellipticity, dB p-p	<6	1
<b>Phase stability:</b>		
Uplink waveguide Δ f/f	$0.5 \times 10^{-15}$ per 1000 s	$0.31 \times 10^{-15}$ per 1000 s <sup>e</sup>
Downlink waveguide Δ f/f	$0.5 \times 10^{-15}$ per 1000 s	$0.17 \times 10^{-15}$ per 1000 s <sup>e</sup>

Notes:

<sup>a</sup>Gain dependent on reflector RMS surface tolerance and other mechanical parameters of antenna structure.

<sup>b</sup>System noise temperature and bandwidth specification dependent on FET bandwidth. A 14 K noise contribution was assumed to achieve this value.

<sup>c</sup>Dependent on achievement of 100-MHz bandwidth and 3.5 K noise contribution.

<sup>d</sup>System noise temperature and bandwidth specification dependent on FET bandwidth. A 28 K noise contribution was assumed to achieve this value.

<sup>e</sup>A warmup period may be required before the Antenna Microwave Subsystem is able to maintain required phase stability.

**Table 2. Antenna Microwave Subsystem Operation System Temperature Noise Estimates**

Element	Noise, K	Element	Noise, K
<b>TWM I low noise</b>		<b>TWM II diplexed</b>	
Feed noise	8.03	Feed noise	21.30
Maser	3.5	Maser	3.5
RCV followup	0.5	RCV followup	0.5
Atmosphere	2.5	Atmosphere	2.5
Galactic	2.7	Galactic	2.7
System temperature (ground-zenith)	17.23	System temperature (ground-zenith)	30.50
Antenna noise	3.3	Antenna noise	3.3
Total system noise (antenna-zenith)	20.53	Total system noise (antenna-zenith)	33.80
<b>TWM II low noise</b>		<b>X-band FET</b>	
Feed noise	11.96	Feed noise	18.97
Maser	3.5	FET	28.0
RCV followup	0.5	RCV followup	0.5
Atmosphere	2.5	Atmosphere	2.5
Galactic	2.7	Galactic	2.7
System temperature (ground-zenith)	21.16	System temperature (ground-zenith)	52.67
Antenna noise	3.3	Antenna noise	3.3
Total system noise (antenna-zenith)	24.45	Total system noise (antenna-zenith)	55.97
<b>TWM I diplexed</b>		<b>S-band FET</b>	
Feed noise	17.38	Feed noise	19.18
Maser	3.5	FET	14.0
RCV followup	0.5	RCV followup	0.5
Atmosphere	2.5	Atmosphere	2.5
Galactic	2.7	Galactic	2.7
System temperature (ground-zenith)	26.58	System temperature (ground-zenith)	38.88
Antenna noise	3.3	Antenna noise	10.60
Total system noise (antenna-zenith)	29.88	Total system noise (antenna-zenith)	49.48

ORIGINAL PAGE IS  
OF POOR QUALITY.

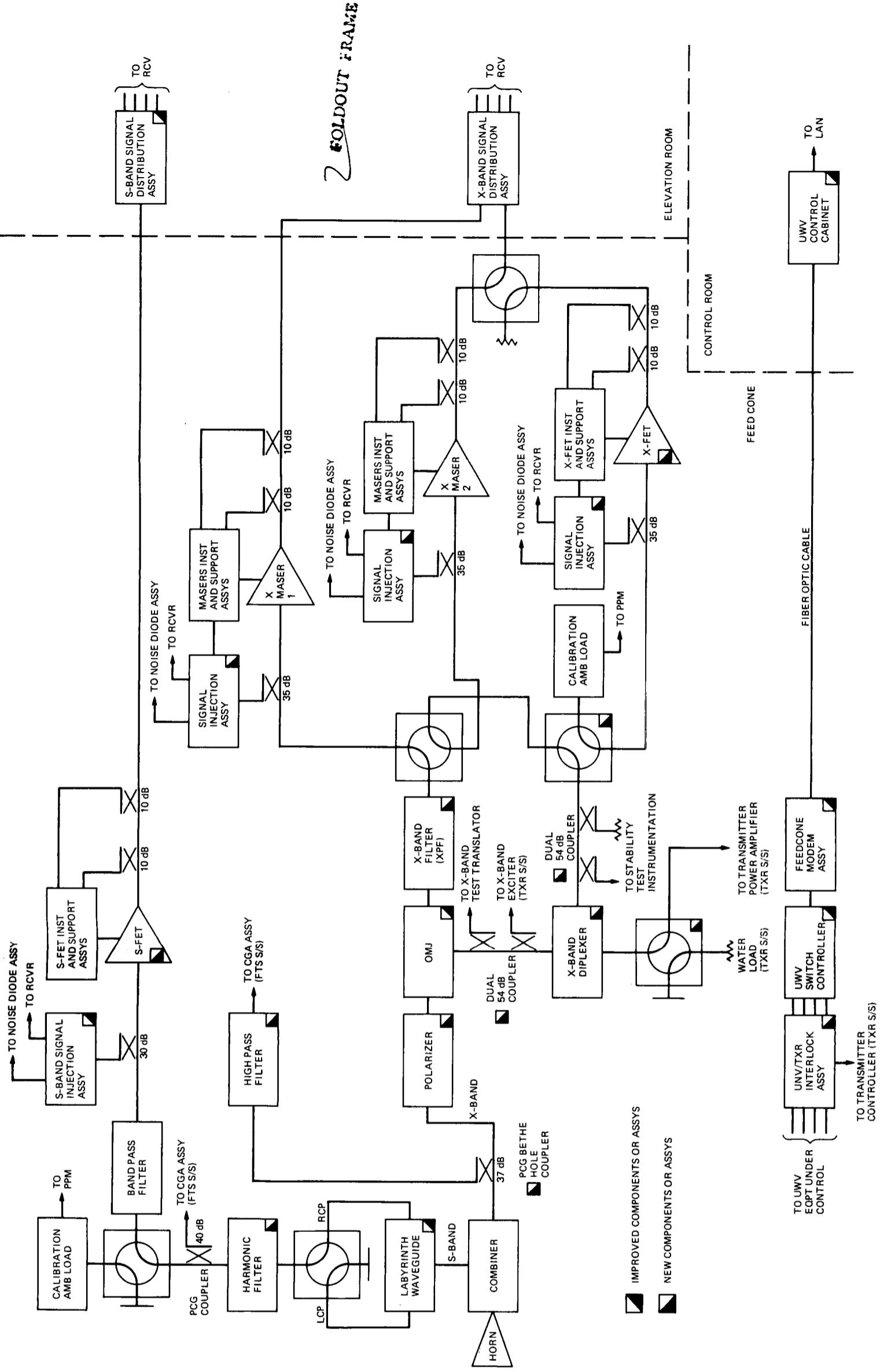


Fig. 1. XSC Functional Block Diagram

FOLDOUT FRAME

PRECEDING PAGE BLANK NOT FILMED

# A New Algorithm for Microwave Delay Estimation From Water Vapor Radiometer Data

S. E. Robinson  
Tracking Systems and Applications Section

*A new algorithm has been developed for the estimation of tropospheric microwave path delays from water vapor radiometer (WVR) data, which does not require site and weather dependent empirical parameters to produce high accuracy. Instead of taking the conventional linear approach, the new algorithm first uses the observables with an emission model to determine an approximate form of the vertical water vapor distribution which is then explicitly integrated to estimate wet path delays, in a second step. The intrinsic accuracy of this algorithm has been examined for two channel WVR data using path delays and simulated observables computed from archived radiosonde data. It is found that annual RMS errors for a wide range of sites are in the range from 1.3 mm to 2.3 mm, in the absence of clouds. This is comparable to the best overall accuracy obtainable from conventional linear algorithms, which must be tailored to site and weather conditions using large radiosonde data bases. The accuracy and flexibility of the new algorithm are indications that it may be a good candidate for almost all WVR data interpretation.*

## I. Introduction

Very Long Baseline Interferometry (VLBI) and Global Positioning System (GPS) applications, as well as other fields, require calibration of microwave delays caused by water vapor in the troposphere. Two and three channel water vapor radiometers (WVRs) are increasingly seen as the best solution to this problem. Unfortunately, WVRs do not directly measure delay; they actually measure antenna temperatures in frequency channels close to the 22.2 GHz resonance of water. The WVR observables must then be related to the desired wet path delays by a data interpretation algorithm. In order to

most effectively use the WVR data, one must, therefore, understand and quantify the level of error in the delay estimation contributed by the algorithm itself.

The most popular algorithm formulation, resulting from the conventional approach to delay estimation, assumes that the microwave delay may be written as a linear expansion in the total optical depths of the troposphere at the WVR channel frequencies (e.g., Refs. 1, 3, and 6). The optical depth at any point  $s$ , for a frequency  $\nu_i$ , is defined by the integral along the line-of-sight in the atmosphere,

$$\tau(s, \nu_i) = \int_0^s ds' \alpha(s', \nu_i)$$

The total optical depth to be used in the delay expansion is, therefore, just  $\tau(\infty, \nu_i)$ . The symbol  $\alpha$  represents the extinction coefficient, also called opacity (e.g., Ref. 5, page 23), a complicated function of temperature, pressure, frequency, and the densities of water vapor and liquid, that is defined by an atmospheric emission model (e.g., Refs. 9 and 10). In the linear algorithms, the  $\tau(\infty, \nu_i)$  are usually estimated from the antenna temperatures after assuming effective radiating temperatures for the tropospheric water vapor. The coefficients in the linear delay expansion, sometimes called retrieval coefficients (see Ref. 3), are determined empirically, usually using a set of wet delays and corresponding simulated WVR antenna temperatures computed from archived radiosonde data. Once the delay coefficients have been determined by regression analysis of the simulated observables, the WVR user has a very simple, easily applied, expression for wet delay.

There are also undesirable features of linear delay algorithms, as with most other algorithms. In particular, the linear coefficients must be empirically tailored to each observing site and set of observing conditions in a process sometimes called stratification, in order to achieve high delay estimation accuracy (Ref. 3). The need for this arises from the fact that the WVR antenna temperatures and surface meteorology data do not uniquely define the implicit parameters of the delay estimation problem. Specifically, the WVR antenna temperatures,  $T_a(\nu_i)$ , are imperfect measurements of the radiation brightness temperatures, which are given by

$$T_b(\nu_i) = \int_0^{\tau(\infty, \nu_i)} d\tau T_k(\tau) \exp(-\tau) + T_{bg} \exp(-\tau(\infty, \nu_i))$$

or, equivalently,

$$T_b(\nu_i) = \int_0^\infty ds \alpha(s, \nu_i) T_k(s) \exp(-\tau(s, \nu_i)) + T_{bg} \exp(-\tau(\infty, \nu_i))$$

where  $\tau$  is defined above,  $T_k$  is the kinetic temperature, and  $T_{bg}$  is the background temperature. Because of the dependence on  $\tau$  and  $\alpha$ , the WVR observables also depend on the distributions of temperature, pressure, and water liquid, as well as the actual amount of water vapor along the line-of-sight. The conventionally defined wet path delay has a simpler form,

$$R_{wet} \propto \int_0^\infty ds \rho_v(s) / T_k(s)$$

but also depends on the distribution of water vapor density,  $\rho_v(s)$ , and the temperature profile (e.g., Ref. 2). One can see from these expressions that only a knowledge of the line-of-sight temperature, pressure, water vapor, and water liquid distributions can uniquely specify both brightness temperatures and wet path delay. Since this information is not directly available in observable quantities, the past values of the required data under specific conditions are used to improve the definitions of the linear delay coefficients.

Stratification of the coefficients requires that a large radiosonde data base be available for the empirical estimation of each coefficient set. The first difficulty that this presents is that of simply matching coefficients to individual sites. Since radiosondes are flown at relatively few locations which are usually near population centers and away from most radio observatories, there is nearly always some uncertainty in the association of a data base with any specific site of interest. In addition, even with the currently discussed computation of delay coefficients for specific seasons, weather conditions, and sites, use of the linear algorithm may become awkward when processing typical VLBI and GPS experiments. Finally, the delay estimation accuracy of the linear algorithms cannot be improved indefinitely by simply increasing the level of stratification, because of the finite number of meaningful observable criteria for discrimination.

In order to avoid the disadvantages of linear algorithms, a new algorithm, the profile algorithm, has been developed from an entirely different approach. The new algorithm's formulation of the delay estimation problem relies on the direct application of the known physics of the problem to make better use of observables, rather than on large data bases to adjust linear delay coefficients for specific conditions. The result is more complex than the linear algorithms, but has considerable advantages. Most notably, it offers complete independence from site and condition-specific empirical parameters, and their inherent problems, while producing delay estimation accuracy that is as good as that of the better stratified linear algorithms.

## II. Description of the Profile Algorithm

The basic premise of the profile algorithm is that an actual solution to a simplified version of the nonlinear delay problem can be generated numerically from each WVR observation and the corresponding surface meteorology. This leads to a formulation and an execution very unlike those for the linear algorithms. In fact, the profile algorithm makes no attempt to impose a direct relationship, linear or otherwise, between WVR antenna temperatures and wet path delays. Instead, simple vertical distributions of water vapor and liquid are

adjusted until they produce, through an atmospheric emission model (e.g., Ref. 10), brightness temperatures in agreement with the WVR antenna temperatures. The delay estimates themselves are produced in a second step by integrating along the line-of-sight using the fitted water vapor distributions.

It is assumed that the vertical distribution of the relative humidity can be adequately represented by a two piece linear function. At surface level, the relative humidity is naturally given by surface meteorology, and at altitude of 10 km, so little water vapor can exist that it can safely be taken as zero. The relative humidity profile is then approximated by connecting the two end points linearly to the relative humidity at an intermediate altitude of 3 km. It is the value of the relative humidity at this point that the algorithm freely adjusts in order to completely specify the vapor profile. An example of such a simplified water vapor profile used by the profile algorithm is illustrated on the left side of Fig. 1. The corresponding water vapor density profile, defined by the sample relative humidity distribution, is shown in the right portion of the figure.

The vertical pressure and temperature profiles must also be known, or assumed, in order to specify the water vapor density and opacity from the algorithm's relative humidity profile. The pressure can be well represented by an exponential decay from the surface value. The temperature profile for the profile algorithm is produced by first creating a nominal temperature profile from data in Tables 5.1 and 5.2 of the Standard Atmosphere (Ref. 12), interpolated to the latitude and altitude of the site, and to the time of year of the observation. The surface temperature implied by the nominal profile is then subtracted from the measured surface temperature and multiplied by  $\exp(-h/H)$ , where  $h$  is the altitude, and  $H$  is a constant scale height, to produce an altitude-dependent temperature correction. The temperature profile used by the algorithm is then formed by adding the nominal profile and the correction. It, therefore, matches the measured surface temperature, and exponentially approaches the nominal temperature distribution as the altitude increases. An examination of temperature profiles from radiosondes at a number of locations suggests that the best choice of scale height for the decay of the surface temperature adjustment is about 2 km. An example of such an adjusted profile is shown in Fig. 2. Note that the site altitude and latitude, needed to produce the nominal temperature profiles, are the only site-specific parameters required by the profile algorithm.

In reality, the actual vertical water vapor distributions deviate from the assumed form of the profiles; however, the impact of these differences is quite small, as the next section of this article demonstrates. One reason for this is that the channel frequencies of the WVRs are usually selected so as to

minimize the influence of changes in the vertical water vapor profiles on estimated delay (Refs. 7 and 11). Another is that in integrating the vapor distribution to find the wet delay, high frequency spatial variations tend to average out, making a smooth function an adequate representation. Still, dependence on the vertical structure cannot be completely eliminated, so the fact that the new algorithm provides even an approximate adjustment to the effective height of the water vapor gives some improvement over the linear algorithms.

In the simulations that are described in the next section, it is assumed that the liquid opacity is negligible, but in actual use the delay estimation algorithm must accurately account for the continuum emission from water droplets in clouds in order to produce high quality delay estimates. The liquid opacity is not a function of pressure, and only weakly dependent on temperature (e.g., Ref. 9), which means that the liquid contribution to the WVR antenna temperatures is relatively insensitive to the actual form of the vertical distribution of droplets. Therefore, the simple assumption that the water liquid density is proportional to the saturation water vapor density at any point has been made, and the algorithm is allowed to fit for the constant of proportionality. The effect of this assumption will be examined in work, unfinished at the time of this writing, dealing with algorithm performance in the presence of clouds.

### III. Accuracy of the Profile Algorithm

All algorithms for the estimation of wet path delays from WVR data contribute some error to the final estimates. This error is intrinsic to the algorithm, and as such can be separated from other error sources. The best way to isolate and quantify the algorithm error is by generating simulated WVR antenna temperatures and corresponding wet path delays from a set of vertical profiles of water density, temperature, and pressure. This is a process similar to that used to generate empirical coefficients for linear algorithms, as described by Gary et al. (Ref. 3), Resch (Ref. 6), and others. Such a simulation has been done for the new profile algorithm, and the results are described in this section.

Radiosonde data are taken on a regular basis at a large number of sites over the world, and can be used as the basis of the required simulation. These data typically consist of temperature and relative humidity on a vertical grid of pressure, which can be converted to altitude. For our simulation, data were obtained spanning a full year for 16 sites of interest. Some sites were selected from the available radiosonde archive because they were closest to prominent VLBI observatories, others were chosen to overlap with other delay retrieval accuracy studies, and the rest were picked to produce a good cross-section of locations in the continental United States.

Although 13 of the data sets contained two radiosonde profiles for most days, Edwards AFB, Madrid, and Wagga-Wagga had less than half that number available. These were, nonetheless, included because of their proximity to Goldstone, DSS 63, and DSS 43, respectively. Radiosonde profiles which had missing or blatantly erroneous data below 5 km, or which terminated at anomalously low altitudes, were removed by software to avoid corruption of the results. This left an average of 645 profiles for all but the three previously mentioned sites, which averaged 260. Examination of the rejected data revealed no selection effects that could bias the simulation.

Next, each profile was used to compute a zenith delay, due to water vapor, and the corresponding antenna temperatures that would be observed by an ideal noise-free WVR with frequency channels at 20.7 GHz and 31.4 GHz. This particular WVR configuration was selected because it corresponds to the majority of existing water vapor radiometers. The emission model used to calculate the WVR observables is essentially that of Waters (Ref. 10) with oxygen opacities given by Rozenkranz (Ref. 8). Because the same emission model was used for both the simulation and the delay retrieval, all emission model errors should cancel in the simulation. Of course, these errors will contribute to the ultimate delay estimation accuracy, but they are the same for all algorithms. No cloud contribution to the continuum emission was used, at this time, because no information concerning liquid content is present in the radiosonde profiles, and a suitable cloud model had not yet been implemented.

Finally, the simulated WVR antenna temperatures and surface meteorology data were used as input to the profile algorithm to produce the delay estimates. The level of algorithm error is then given by the difference between the retrieved delay estimates and the corresponding delays computed directly from the radiosonde profiles. These results for all 16 sites appear in Table 1. The last column displays the root-mean-square algorithm error for the entire year at each location. These range from 1.3 mm for Wagga-Wagga up to 2.3 mm for Munich, and have an average value of 1.8 mm. The annual average zenith delay and the maximum error in estimated delay are also displayed in the table.

Resch presented several variations on the linear delay retrieval algorithm in his 1983 paper (Ref. 6). The most popular, and most accurate, makes use of surface meteorology, and fits the standard form of the linear algorithm given below.

$$R_{\text{wet}} = c_0 / \mu + a_0 + a_1 \tau(\infty, \nu_1) + a_2 \tau(\infty, \nu_2)$$

$R_{\text{wet}}$  is the wet path delay estimate, the  $a_i$  and  $c_0$  are empirically determined constants, and  $\mu$  is the cosine of the zenith

angle of the observation. The  $\tau(\infty, \nu_i)$  are optical depth estimates, corresponding to the WVR frequency channels, computed from

$$\tau(\infty, \nu_i) = -\ln [(T_{\text{eff}}(\nu_i) - T_a(\nu_i)) / (T_{\text{eff}}(\nu_i) - 2.9)]$$

The  $T_{\text{eff}}(\nu_i)$  in the above equation are empirically determined effective radiating temperatures, computed by fitting to simulated WVR observations, and the  $T_a(\nu_i)$  are the measured antenna temperatures.

For this algorithm, Resch determined best-fits to site-specific linear delay coefficients using 90 radiosonde profiles to cover a one-year period for each of five different locations. He then presented the root-mean-square errors of those fits to the data bases under the same assumptions of no liquid opacity and no WVR noise that were used in the simulations here. The results in this article may, therefore, be compared directly to his for these five sites, as has been done in Table 2. Although the errors for the individual sites are generally different, the average RMS error, at 2 mm, is not significantly different for the stratified linear algorithm and the profile algorithm with no empirically determined site-specific parameters, at all.

#### IV. Discussion and Conclusions

Because of its need to perform emission model calculations, the profile algorithm is required to do a relatively large amount of computation in order to estimate delays from WVR observations. Linear algorithms, on the other hand, require extensive calculations to initially produce linear delay coefficients, but thereafter need very little computation to produce delay estimates. This is not a serious drawback for the new algorithm, however, since even the microcomputers that run the more recent WVRs, such as Janssen's J01 (Ref. 4), are fully capable of running real-time software implementing the profile algorithm. The emission model calculations also produce a definite advantage in that the tropospheric emission characteristics are imperfectly known, and subject to revision. Any emission model changes made to the profile algorithm code can immediately be used for reducing data without delay. In contrast, linear algorithms require a repetition of the lengthy simulations, reprocessing all of the radiosonde data originally used to generate the linear delay coefficients, before new delay estimates may be produced.

There may well be instances when WVR users would prefer to spend only the smaller amount of computation time needed by the linear algorithms, without accepting their less desirable features. For such situations, suitable hybrid schemes can be devised. For example, the profile algorithm is particularly well-suited for use with tipping-curve reduction programs; a

very simple extension would allow it to use the tipping-curve data to simultaneously produce linear delay coefficients, which could then be used for delay estimation until the next invocation of the profile algorithm. There are, of course, many other possible ways of making the best use of both types of algorithms in combination.

The work presented here demonstrates the validity of the profile algorithm and, since many other related formulations are possible, the approach to the WVR delay estimation problem that the new algorithm represents. The most serious question still to be addressed is the profile algorithm accuracy

in the presence of clouds. Preliminary indications from work now in progress are that, while cloudy weather does decrease the delay estimation accuracy, the overall algorithm accuracy is not degraded by more than 15 percent by the addition of liquid opacity to the simulations. Even without the liquid opacity information, the flexibility of the new algorithm, and its apparent ability to use the same procedures and parameters to reduce WVR data from any site under a wide range of weather conditions, while still obtaining high delay estimation accuracy, may make it a good candidate for use with almost all WVR applications. Additional work to thoroughly investigate its properties is certainly justified.

## Acknowledgment

We would like to thank Don Trask for his helpful discussions and encouragement in developing this algorithm.

## References

1. Claflin, E. S., Wu, S. C., and Resch, G. M., "Microwave Radiometer Measurement of Water Vapor Path Delay: Data Reduction Techniques," *DSN Progress Report 42-48*, pp. 22-30, Jet Propulsion Laboratory, Pasadena, CA, Dec. 15, 1978.
2. Crane, R. K., "Refraction Effects in the Neutral Atmosphere," *Methods of Experimental Physics: Astrophysics*, M. L. Meeks, ed., Vol. 12, Part B, pp. 186-200, 1976.
3. Gary, B. L., Keihm, S. J., and Janssen, M. A., "Optimum Strategies and Performance for the Remote Sensing of Path-delay Using Ground-based Microwave Radiometers," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. GE-23, No. 4, pp. 479-484, July 1985.
4. Janssen, M. A., "A New Instrument for the Determination of Radio Path Delay Due to Atmospheric Water Vapor," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. GE-23, No. 4, pp. 485-490, July 1985.
5. Mihalas, D., *Stellar Atmospheres*, W. H. Freeman and Company, San Francisco, CA, p. 23, 1978.
6. Resch, G. M., "Inversion Algorithms for Water Vapor Radiometers Operating at 20.7 and 31.4 GHz," *DSN Progress Report 42-76*, pp. 12-26, Jet Propulsion Laboratory, Pasadena, CA, Feb. 15, 1984.
7. Resch, G. M., "Another Look at Optimum Frequencies for a Water Vapor Radiometer," *DSN Progress Report 42-76*, pp. 1-11, Jet Propulsion Laboratory, Pasadena, CA, Feb. 15, 1984.

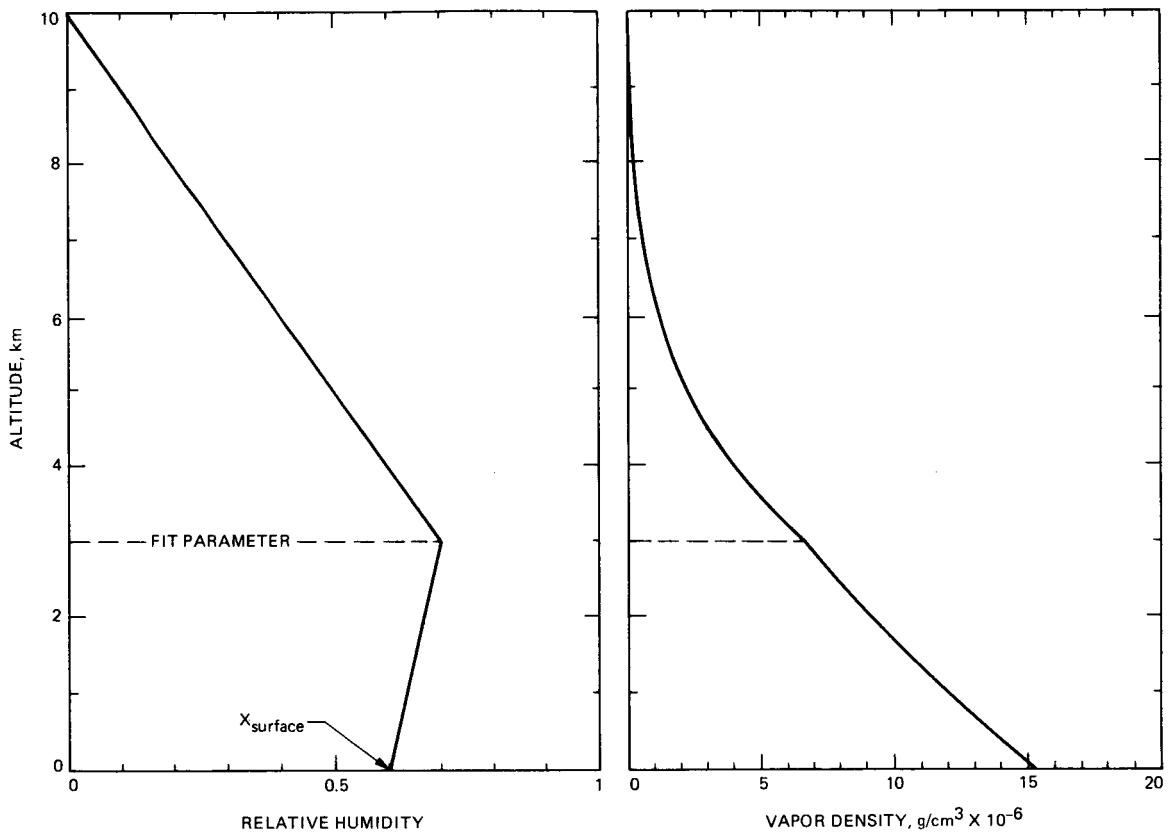
8. Rosenkranz, P. W., "Shape of the 5 mm Oxygen Band in the Atmosphere," *IEEE Transactions on Antennas and Propagation*, Vol. AP-23, No. 4, pp. 498-506, July 1975.
9. Staelin, D. H., "Measurements and Interpretations of the Microwave Spectrum of the Terrestrial Atmosphere near 1-cm Wavelength," *Journal of Geophysical Research*, Vol. 71, pp. 2875-2882, 1966.
10. Waters, J. W., "Absorption and Emission by Atmospheric Gases," *Methods of Experimental Physics: Astrophysics*, M. L. Meeks, ed., Vol. 12, Part B, pp. 142-175, 1976.
11. Wu, S. C., "Optimum Frequencies of a Passive Microwave Radiometer for Tropospheric Path-Length Correction," *IEEE Transactions on Antennas and Propagation*, Vol. AP-27, No. 2, pp. 233-239, Mar. 1979.
12. *U. S. Standard Atmosphere Supplements, 1966*, Tables 5.1-5.2, pp. 99-203, prepared under sponsorship of ESSA, NASA, and the USAF, U. S. Government Printing Office, Washington, DC, 1966.

**Table 1. Profile algorithm errors for 1979**

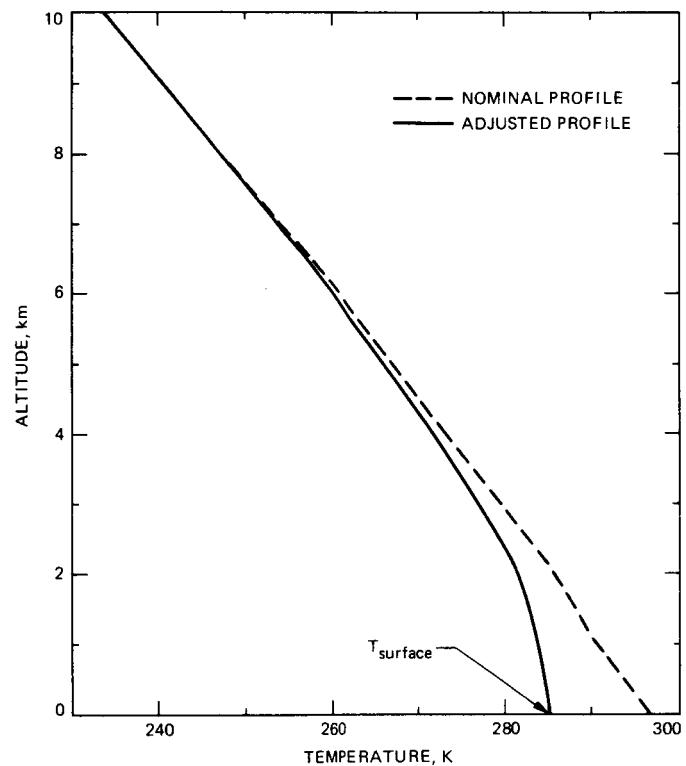
Site	Average delay, cm	Maximum error, cm	RMS error, cm
Albuquerque	7.4	0.55	0.17
Apalachicola	18.0	0.85	0.19
Boise	7.4	0.52	0.16
Dayton	12.3	0.90	0.20
Denver	6.6	0.62	0.14
Edwards AFB	6.6	0.51	0.17
El Paso	8.7	0.92	0.17
Goteborg	8.0	0.77	0.16
Madrid	10.4	0.95	0.19
Munchen	10.0	0.81	0.23
Oakland	9.0	0.64	0.20
Oklahoma City	12.5	0.93	0.20
Pittsburgh	11.0	0.77	0.18
Portland (ME)	10.3	0.97	0.22
San Diego	9.0	0.84	0.22
Wagga-Wagga	9.1	0.41	0.13
	Minimum RMS error	0.13	
	Maximum RMS error	0.23	
	Average RMS error	0.18	

**Table 2. Comparison of errors from profile and Resch algorithms**

Site	Average delay, cm	Profile RMS error, cm	Resch RMS error, cm
El Paso	8.7	0.17	0.14
Oakland	9.0	0.20	0.27
Pittsburgh	11.0	0.18	0.16
Portland (ME)	10.3	0.22	0.18
San Diego	9.0	0.22	0.30
	Minimum RMS error	0.17	0.14
	Maximum RMS error	0.22	0.30
	Average RMS error	0.20	0.21



**Fig. 1.** An example of the assumed relative humidity profile appears on the left. The dashed line at 3 km represents the value of the fitting parameter. The water vapor density profile corresponding to the relative humidity is shown on the right.



**Fig. 2.** An example of a nominal temperature profile derived from data in the Standard Atmosphere is represented by the dashed line. The solid line shows the temperature profile after it has been forced to match the measured surface temperature.

D17-36

N87-15346

TDA Progress Report 42-87

July - September 1986

## Block IIA Traveling-Wave Maser

D. L. Trowbridge

Radio Frequency and Microwave Subsystems Section

Two 8.4-GHz low-noise traveling-wave masers (TWMs) with effective input noise temperatures of 3.6 to 3.9 K and bandwidth in excess of 100 MHz have been supplied to the Deep Space Network. These TWMs are used on the 64-meter antennas at Deep Space Stations 14 and 43 to meet the requirements of the Voyager Uranus encounter. The TWMs have improved isolator assemblies and new interstage matching configurations to reduce gain/bandwidth ripple. They are equipped with followup Field Effect Transistor Amplifiers as part of the design to meet the 100-MHz bandwidth requirements of very long baseline interferometry.

### I. Introduction

Two new 8.4-GHz low-noise traveling-wave masers (TWMs) have been installed on the 64-meter antennas at Deep Space Stations 14 and 43 as part of the Mark IVA DSCL Antenna Microwave Subsystem upgrade. The TWMs were built to meet the sensitivity requirements of the Voyager Uranus encounter and support very long baseline interferometry (VLBI) bandwidth requirements.

A previous version of the maser, designated Block II TWM (Ref. 1), met the noise requirements for the Voyager Saturn encounter but failed to meet the new 100-MHz bandwidth requirements. A new prototype TWM was designed and built that would meet requirements of both the Voyager Uranus encounter and VLBI. This new 8.4-GHz TWM is identified as the Block IIA Low-Noise TWM and is shown in Figs. 1, 2, and 3. This maser provides 45-dB net gain (measured between the input of the TWM and the input to the receiver mixer), with a -3-dB bandwidth greater than 100 MHz (see Table 1) and an effective input noise temperature of 3.6 to 3.9 K.

### II. Design Goals

The following performance design goals were established for the Block IIA Low-Noise TWM:

- (1) Gain  $44.5 \pm 1.5$  dB, to be adjusted on the particular Front End Area (FEA) of the antenna for net gain of combined TWM, Field Effect Transistor (FET) Amplifier, and hybrid loss in front of the receiver mixer.
- (2) 100-MHz minimum -3-dB bandwidth covering 8400 to 8500 MHz.
- (3) Equivalent input noise temperature of  $4 \pm 1$  K within the specified bandwidth.
- (4) Gain slope within the specified bandwidth  $\pm 0.2$  dB/ MHz.
- (5) Improved pump source assembly reliability and ability to meet Radio Frequency and Microwave Subsystems Section (Section 333) stress analysis standards.

- (6) Closed-cycle refrigeration (CCR) stage temperature monitoring capability.
- (7) Interchangeability with existing Block II TWMs (Ref. 1) and previous Block I TWMs (Ref. 2).

Table 1 lists additional design requirements for the Block IIA TWM.

### III. Maser Description

The Block IIA Low-Noise TWM external package is identical to the Block II TWM (Ref. 1) with the exception of the FET amplifier assembly, mounted near the output waveguide port of the TWM. Internal cryogenic temperature sensors are mounted on the three temperature stages of the CCR (70 K, 15 K, and 4.5 K). This allows the stage temperature to be monitored during cool-down and when the unit is cold. Monitoring of these sensors is useful in troubleshooting the CCR system during cool-downs and while the unit is cold. A new magnet field monitor sensor results in more accurate and repeatable magnetic field monitoring.

The major performance design changes are within the maser structure assembly. The previous DSN Block II 8.4-GHz TWMs had marginal gain bandwidth product. Most of the six units built could not meet the minimum gain requirement of 45-dB gain with 100-MHz bandwidth. The gain was adjusted for 45-dB gain and the resulting bandwidth was less than 100 MHz on most units (actual bandwidth realized varied from 65 MHz to 108 MHz on the six Block II TWMs). The cause of low gain/bandwidth product was inadequate inversion ratio and excessive ripple in the gain/bandwidth response.

Low inversion ratio results when an insufficient number of maser spins make the energy level transition upon application of pump energy. Attempts to analyze the pump frequency RF fields in the TWM structure have met with little success due to the complexity of the dielectric loading, presence of the half-wave comb slow-wave structure, and overmoding of the pump energy within the structure. Measured variations in inversion ratio have resulted from changes in comb resonant strip material thickness, comb resonant material (pure copper vs. thin-film copper over chrome), and dielectric material shape. Hand-assembled combs using pure copper resonator strips consistently produced better inversion ratios than those obtained with thin electro deposit film strips. Also, attempts to increase the gain bandwidth product of the structure by changing the ruby geometry (a bevel at the edge of the ruby opposite the resonant fingers increases the slowing factor) resulted in lower inversion ratios. Therefore, maximum inversion ratio in this Block IIA TWM was achieved by using hand-fabricated copper combs with minimum ruby shaping.

The geometry of the dielectric-loaded pump coupling waveguide was also modified to improve the inversion ratio.

Figure 4 shows a Block II TWM gain bandwidth curve, selected to show a worst-case sample of gain ripple. This ripple is caused by a combination of insufficient isolator reverse loss and interchannel voltage standing wave ratio (VSWR). Reflected signal due to marginal VSWR and insufficient isolator reverse loss results in in-phase and out-of-phase signal voltage additions, causing peaks and nulls in the maser bandpass. The electrical length of the forward and return path of each structure determines the frequency difference in MHz between adjacent peaks and nulls. This condition was improved in this block IIA design by improving the impedance match at the input and output connector of each channel, and by increasing reverse loss of the distributed resonance isolator in each channel. The Block II TWM isolator had nominally 30-dB reverse loss per channel, or 120-dB total reverse loss per TWM consisting of four channels. The new Block IIA isolator assemblies have a reverse loss of 40 dB per channel (160 dB per TWM). The increase in reverse loss is achieved by increasing the quantity of yttrium iron garnet isolator material (YIG) in the area where the RF signal magnetic field is circularly polarized. Increasing the YIG in this area also results in increased forward loss, which is undesirable. The resulting increase in noise temperature contribution from this loss is, however, offset by the reduction in noise temperature due to the use of a followup FET amplifier (discussed later).

The geometry of the Block IIA TWM input, output, and interstage RF coupling probes was redesigned to improve the long-term mechanical stability of these components. This results in a more repeatable VSWR from cool-down to cool-down and after disassembly and reassembly.

The improvements in isolator performance and VSWR stability have resulted in the gain bandwidth curve for the Block IIA TWM shown in Fig. 5. The performance of a second unit is shown in Fig. 6.

An FET amplifier and variable attenuator are used on the Block IIA TWM, shown in Fig. 7, to offset the high-noise contribution which would otherwise result from the followup receiver. The gain of the TWM/FET/attenuator system is set at 52 dB, on the particular 64-meter FEA, by adjusting the variable attenuator. With the 7-dB insertion loss resulting from power dividers and waveguide loss, a net gain of 45 dB results from TWM input to receiver mixer input.

The receiver mixer is specified at 11 dB maximum noise figure and the insertion loss between TWM and receiver is 7 dB maximum, making the total noise figure for the receiver 18 dB.

The followup receiver noise temperature contribution to system noise temperature is given by the expression:

$$T_f = \frac{T_{\text{receiver}}}{G_{\text{maser}}}$$

Therefore, the receiver followup contribution ( $T_f$ ) to total system temperature ( $T_{\text{op}}$ ) in the previous Block II TWM configuration with a gain of 45 dB is 0.57 K. The  $T_f$  contribution to  $T_{\text{op}}$  with the Block IIA TWM plus FET (total gain = 52 dB) is equal to 0.11 K. This is an improvement in  $T_{\text{op}}$  of 0.46 K.

A redesign of the pump source assembly was made in accordance with Section 333 stress analysis standards. A stress analysis of all components used in the pump system was included with this design effort.

#### IV. Performance

The performance of the Block IIA TWM meets all the design goals and specified requirements shown in Table 1.

The gain/bandwidth curves of the two production Block IIA TWMs are shown in Fig. 5. The gain bandwidth curves are smooth with no evidence of the signal addition ripple that is characteristic of Block II TWMs. The net gain of both Block IIA

TWMs can be adjusted to 44.5-dB gain at the input to the receiver mixer with bandwidths exceeding 100 MHz. Figure 8 shows the bandwidth curve of the Block IIA TWM with the gain varied in 3-dB steps ( $\pm 6$  dB total) with no bandwidth shape change as the result of the gain variation.

The equivalent input noise temperature at the room temperature input waveguide flange is shown in Fig. 9 for both Block IIA TWMs. The noise temperature was measured by attaching a high-quality feedhorn to the waveguide input flange and alternately viewing the "cold" sky and an ambient termination microwave absorber.

#### V. Conclusions

The Block IIA 8.4-GHz TWMs have met or exceeded the performance goals established as part of the Mark IVA DSCC upgrade for the Voyager Uranus encounter and VLBI requirements.

Recently, two additional Block IIA TWMs were built by Eaton Corporation Airborne Instrument Laboratory (AIL) division for the European Space Agency with the assistance of JPL personnel as technical consultants. The successful fabrication of the Block IIA type TWM by a commercial company (AIL) demonstrates the repeatability of this TWM design and the success of the Block IIA TWM program at JPL.

#### References

1. Trowbridge, D. L., "X-Band, Low-Noise, Traveling-Wave Maser," *TDA Progress Report 42-60*, pp. 126-131, Jet Propulsion Laboratory, Pasadena, CA, Dec. 15, 1980.
2. Trowbridge, D. L., "X-Band Traveling Wave Maser Amplifier," *DSN Progress Report 42-28*, pp. 69-70, Jet Propulsion Laboratory, Pasadena, CA, Aug. 15, 1975.

**Table 1. X-band Block IIA TWM and CCR assembly functional characteristics**

Characteristics	Required value
Gain	45 ± 1.5 dB
Gain slope in specified bandwidth	±0.2 dB/MHz
Maser gain stability stationary, short-term	±0.03 dB/10 s, any position
Stationary, long-term	±0.5 dB/12 h, any position
Tilting	±0.5 dB/0.2 deg/s max rate, any position
Bandwidth	
Bandwidth	> 100 MHz (3 dB)
Center frequency	8450 MHz
Noise temperature	4.0 ± 1 K
Phase stability	
10 s	±1 deg max
12 h	±5 deg max
Moving antenna	±10 deg max
Group delay stability	
10 s	±0.1 ns max
12 h	±0.5 ns max
Moving antenna	±1.0 ns max
Group delay variations vs frequency	
Maximum in bandpass	10 ns peak to peak
Maximum slope	0.5 ns/MHz

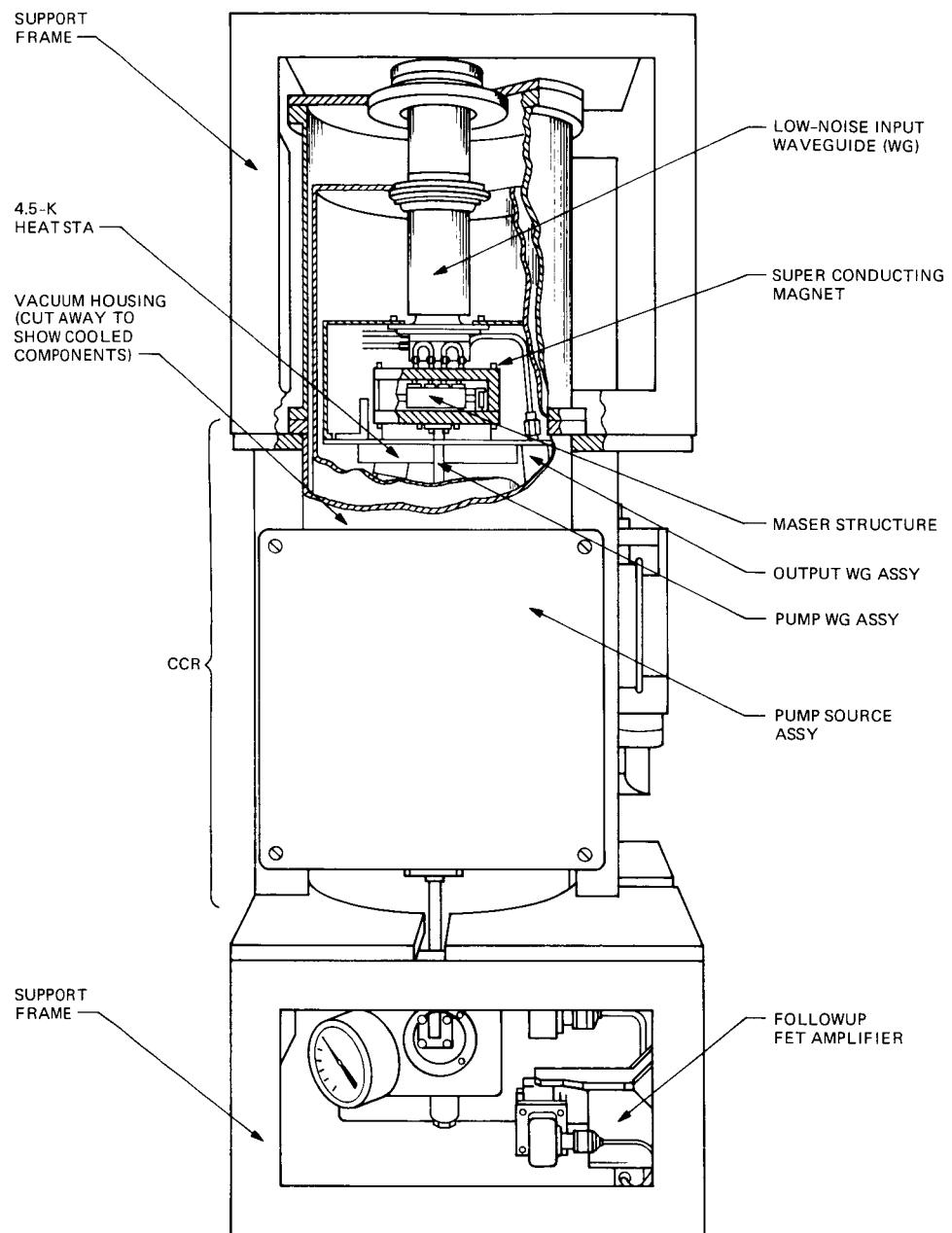


Fig. 1. Traveling-wave maser and closed-cycle refrigerator assembly, Block IIA

ORIGINAL PAGE IS  
OF POOR QUALITY

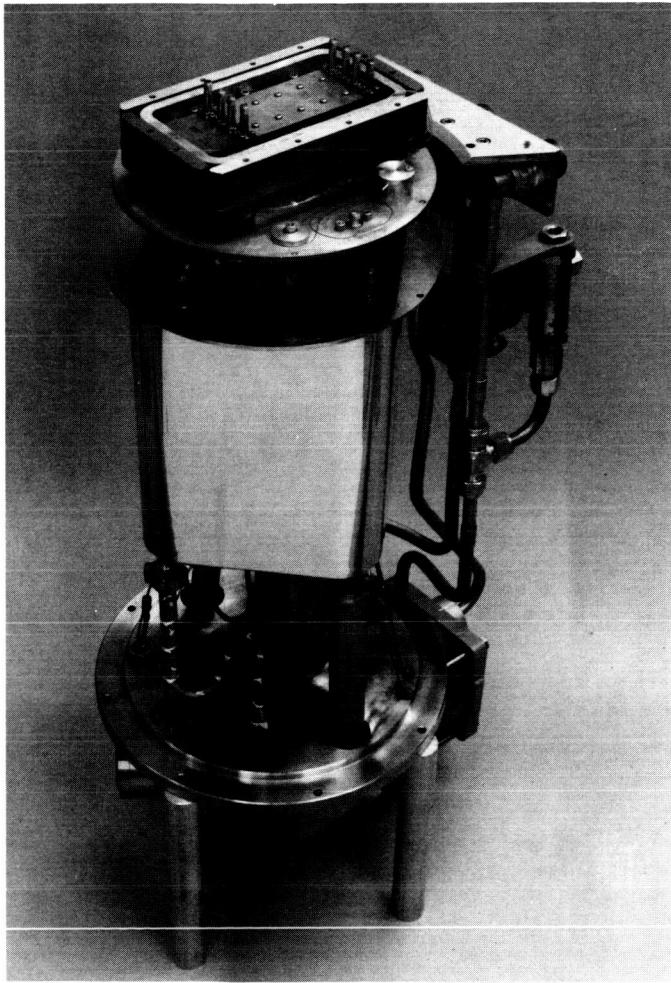


Fig. 2. Internal view of Block IIA 8.4-GHz TWM

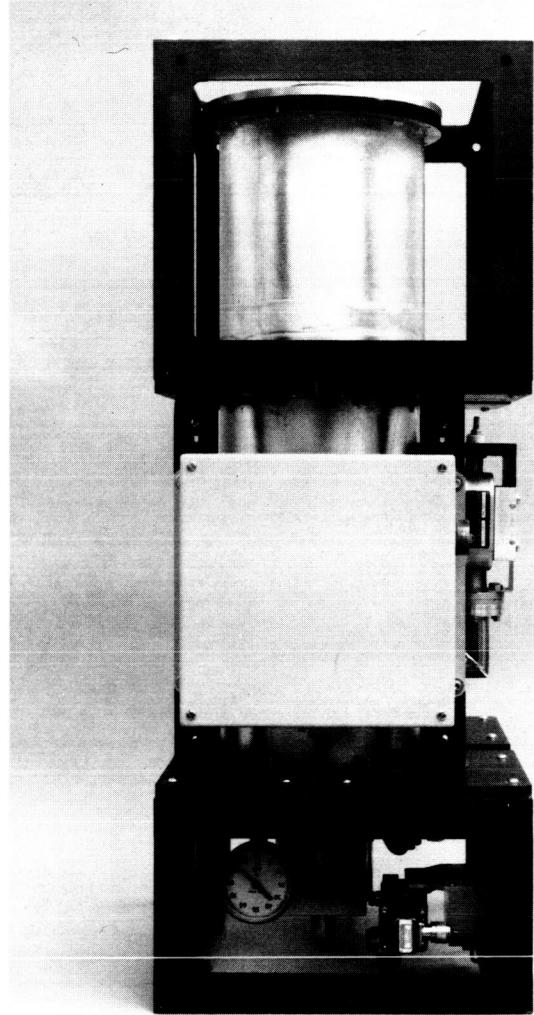


Fig. 3. External view of Block IIA 8.4-GHz TWM

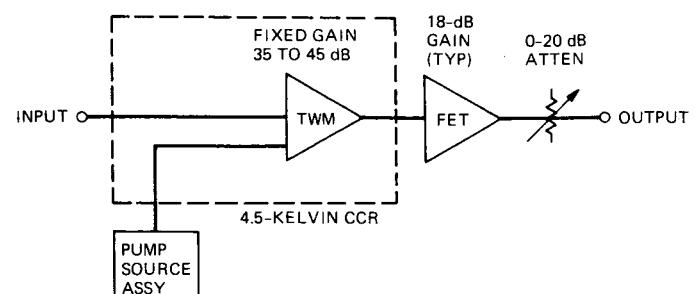
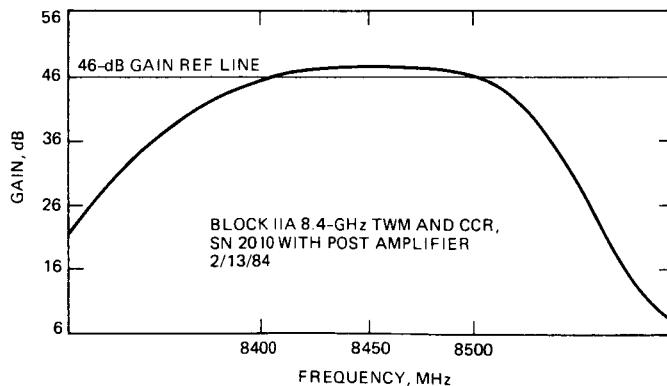
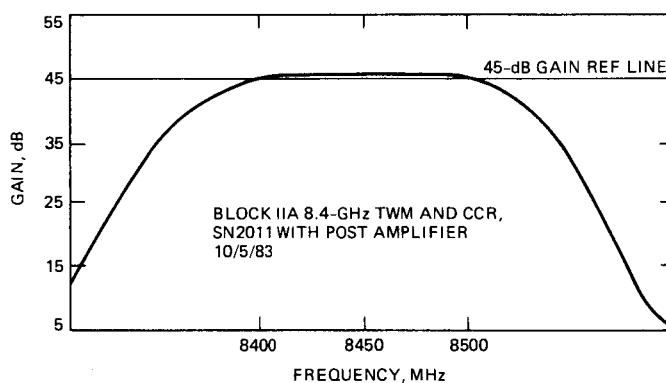
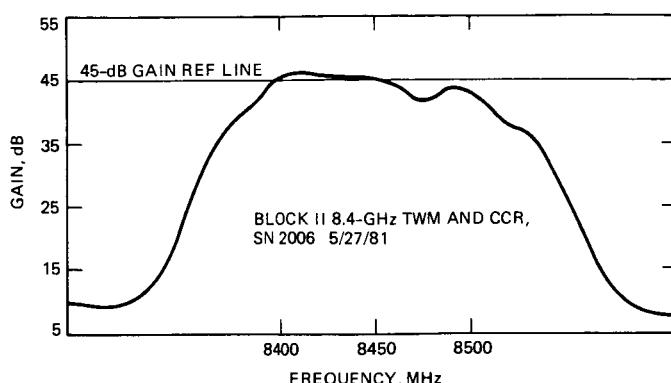
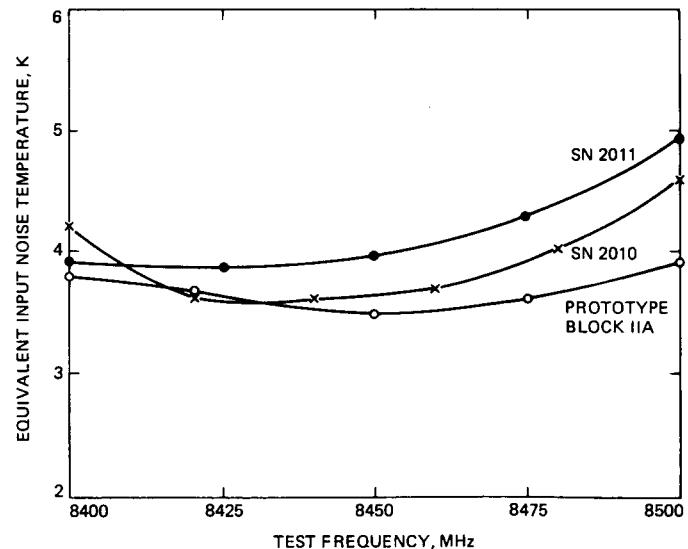
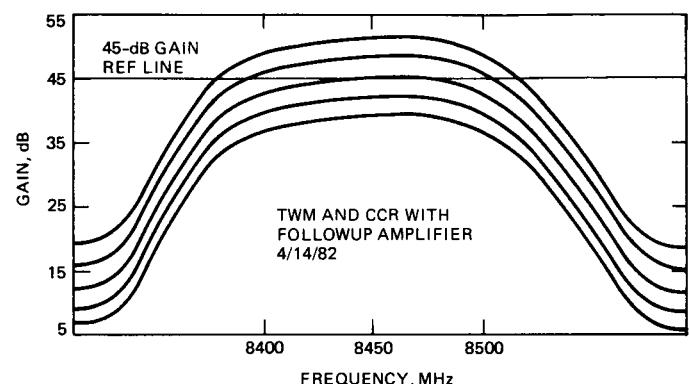


Fig. 7. Block diagram of Block IIA maser



# Antenna Servo Control System Characterization: Rate Loop Analysis for 34-m Antenna at DSS 15

J. A. Nickerson, D. G. Cox, H. K. Smith, J. H. Engel, and H. G. Ahlstrom  
Ground Antenna and Facilities Engineering Section

*This report characterizes the elevation and azimuth servo rate loops at the 34-m High Efficiency Deep Space Station 15 (DSS 15). Time and frequency response performance criteria were measured. The results are compared to theoretically deduced performance criteria. Unexpected anomalies in the frequency response are observed and identified.*

## I. Introduction

This report is the first in a series of reports aimed at studying and characterizing antenna drive control system components of the Deep Space Network antennas. Three types of antennas will be characterized: (a) the high efficiency 34-meter AZ/EL antenna at Deep Space Station 15 (DSS 15), (b) the 34-meter HA/DEC antenna at DSS 12, and (c) the 64/70-meter AZ/EL antenna at DSS 14.

Two major control loops of the antenna control system are the position loop and rate loop. The position loop drives the antenna based on a desired angle command. Figure 1 is a functional block diagram of a typical position control loop. A major component of the position loop is the rate loop. The rate loop is a feedback control system designed to control motor angular velocity. It responds to a motor rate command signal issued from the position loop. Characterization of an antenna's motion control system requires operational knowledge of both the position and rate loops.

Standard control system performance criteria include frequency and time domain characteristics. The frequency response of a system, typically displayed in a Bode plot,

depicts system response in terms of gain and phase lag as a function of frequency. Control system bandwidth is deduced from Bode plots. Typical time domain characteristics describe system dynamics in response to a unit step command. Time domain measurements include rise time, percent overshoot, and settling time (see Ref. 1).

This report presents a frequency response characterization of azimuth and elevation rate loops of the 34-meter antenna at DSS 15. Also, normalized rate loop step responses are presented and discussed. Empirical results are compared to theoretical responses derived from linear mathematical models. Subsequent reports will analyze the position loop for the same antenna and extend control system analysis to other antennas in the Deep Space Network (DSN).

## II. Rate Loop Model

A rate loop is closed around each drive motor using a tachometer as a feedback transducer. The simplified functional block diagram, using Laplace domain models, is presented in Fig. 2. The functional blocks around the simplified loop are: rate loop compensation, a current loop, motor load and

tachometer scaling. The  $K$ 's represent gain and conversion factors,  $\tau$ 's represent time constants of various components, and the  $J$  represents motor inertia. Elevation and azimuth models are identical in form but encompass different coefficient values.

The frequency response of elevation and azimuth rate loops is shown in Figs. 3 and 4, respectively. These frequency responses were generated from the model in Fig. 2 using constants supplied by the manufacturer. The frequency response of the elevation rate loop has a maximum gain of 1.2 dB at 2.2 Hz. Beyond 2.2 Hz the gain rolls off with increasing negative slope until near 20 Hz, the roll off is -40 dB per decade. The gain bandwidth is 9.1 Hz. The phase lag is small for low frequencies and monotonically decreases as a function of frequency. The azimuth rate loop frequency response is similar to the elevation rate loop and has a bandwidth of 9.4 Hz.

The same mathematical models were used to simulate the rate loop time response to a unit step input. Figures 5 and 6 show elevation and azimuth rate loop responses to a unit step. For each axis's rate loop, the rise time and settling time are the same at 0.09 seconds and 0.30 seconds, respectively. The percent overshoot for the elevation loop was 4.0%, which is slightly different from 4.5% for the azimuth axis. Both the low percent overshoot and fast rise time, which the mathematical models suggest, are desirable qualities for a good pointing and tracking control system.

### III. Experimental Procedures

At the DSS 15 antenna site each rate loop was disconnected from the position control computer and connected to the signal generator of a frequency response analyzer. A scaled tachometer feedback voltage was connected to the analyzer. The servo rate loop was driven with a sinusoidal voltage from the signal generator of the frequency analyzer at many frequencies selected to cover the range desired (0.1 – 20 Hz). The amplitude of the tachometer voltage and its phase lag at each frequency were reported by the analyzer and recorded.

In the first set of tests two types of input signals were used to measure the frequency response: (a) a biased sinusoidal input voltage, and (b) an unbiased sinusoidal input voltage. The sinusoidal voltage varied between  $\pm 0.06$  volts (corresponding to an antenna angular velocity of  $\pm 5$  mdeg/s). The bias was 0.12 volts (corresponding to an antenna angular velocity of 10 mdeg/s). Tests were conducted on the elevation rate loop centered around four elevation angles: 7, 45, 70, and 88 degrees. Azimuth rate loop tests were conducted at three elevation angles: 7, 45, and 88 degrees. Only one azimuth angle was tested since the azimuth axis inertia changes only with

elevation angle. The elevation angle was varied during azimuth tests to check for cross coupling and effects due to changing inertia.

The second set of tests was performed which measured the step response of both rate loops. For these tests a portable computer with digital-to-analog (D/A) and analog-to-digital (A/D) conversion capabilities was used. A voltage step was induced into the rate loop using the D/A converter, and the scaled tachometer voltage was recorded by the computer using the A/D converter. The rate loop response was sampled and digitized 50 times/second.

## IV. Results

A total of eight frequency tests was made on both axes: four with biased input signals and four with no bias. A Bode plot was generated for each test as shown in Figs. 5 through 19.

### A. Biased Input Tests — Elevation Loop

Figures 7, 8, 9, and 10 give the results of a biased sinusoidal input voltage for the elevation rate loop. Each was made at four different elevation angles. All four tests had similar results and indicate that the frequency response of the rate loop is not a function of elevation angle. Two abnormalities are seen in the frequency response of Figs. 7 through 10. The first is that a large anti-resonance (inverted resonance) appears in the neighborhood of 2.3 Hz. A second smaller anti-resonance appears near 3.4 Hz. These anti-resonances are seen in both magnitude and phase plots.

The large anti-resonance is believed to be the first mode of torsional vibration between the motor's rotor inertia and the antenna's inertia. This two-degree of freedom system is idealized as two inertias (or masses) connected by a spring. The antenna and the motor rotor are the two inertias while the gear reducer acts as the spring. The inertia of the antenna, as seen by each elevation motor, is reported to be  $0.01898 \text{ kg-m}^2$  ( $0.014 \text{ lb-ft-s}^2$ ). The motor and gearbox inertia, as seen by the motor, is  $0.141 \text{ kg-m}^2$  ( $0.104 \text{ lb-ft-s}^2$ ). The spring constant of the gear reducer is nonlinear and ranges between  $3.96 \text{ N-m/rad}$  ( $1.71 \text{ lb-ft/rad}$ ) and  $5.33 \text{ N-m/rad}$  ( $3.93 \text{ lb-ft/rad}$ ) for torque values less than 25% maximum motor torque. Equation (1) calculates the fundamental natural frequency,  $w_n$ , for a two mass torsional system connected by a spring, where  $K_s$  is the torsional stiffness of the gear reducer,  $J_1$  is the inertia of the motor's rotor, and  $J_2$  is the antenna's inertia as seen by the motor (see Ref. 2):

$$w_n = \sqrt{K_s (J_1 + J_2) / (J_1 J_2)} \quad (1)$$

Using the above values for  $J_1$ ,  $J_2$ , and  $K_s$  suggests the first mode of torsional oscillation is between 1.87 Hz and 2.84 Hz. The large 2.3-Hz anti-resonance frequency is centered between the two frequencies; therefore, it is reasonable to assume that the large anti-resonance is due to this first torsional oscillatory mode.

Note that during first torsional mode vibrations, a motionless node exists on the flexible link between the two inertias. The greater the ratio of inertias the closer the node is to the larger inertia. This indicates that the motion of the larger inertia (which is the motor's rotor) is attenuated. The second small anti-resonance located near 3.4 Hz is also believed to be due to a torsional mode of vibration. This has not, however, been verified.

The frequency responses shown in Figs. 7, 8, 9, and 10 can also be compared with the theoretical frequency response shown in Fig. 3. The theoretical frequency response does not contain the anti-resonances found in the experimental data since these are not accounted for by the model. The actual frequency responses have slightly lower phase and gain margins compared to the theoretical frequency response shown in Fig. 3. The actual bandwidth is 8.8 Hz, which is 4% less than the theoretical bandwidth.

## B. Unbiased Input Tests — Elevation Loop

Figures 11, 12, 13, and 14 show results from the elevation axis rate loop using an unbiased sinusoidal input voltage instead of a biased voltage. Three observations were made from these results: the frequency response is completely different from biased tests yet the same large and small anti-resonances are still observed and the frequency response is not a function of elevation angle. The rate loop bandwidth is reduced from 3.8 Hz to 1 Hz. This is a 90% decrease in bandwidth. Greater phase lag also exists below 3 Hz. One cause of the bandwidth and phase lag behavior could be the susceptibility of the system to friction; the friction occurring when the motors change direction. Changing motor direction reduces rate loop performance.

## C. Biased Input Tests — Azimuth Rate Loop

The tests performed on the elevation axis were repeated on the azimuth axis. The results from biased input are plotted in Figs. 15, 16, and 17. The tests were performed at the three elevation angles of 7, 45, and 88 degrees, respectively. Again, the frequency response was found not to be a function of elevation angle. All three frequency responses are similar; but comparisons between the experimental results and theory show a significant difference. The measured rate loop bandwidth is greater than what theory predicts. A 27% increase is

seen from 9.4 Hz to 12 Hz. Analyzing phase margin yields similar comparisons. A small dip was observed in the gain at 2 Hz in all three Bode plots; otherwise, no large anti-resonances were observed. The first mode is calculated from the azimuth axis inertias,  $J_1 = 0.123 \text{ kg-m}^2$  ( $0.0907 \text{ lbf-ft-s}^2$ ) and  $J_2 = 0.0118 \text{ kg-m}^2$  ( $0.0087 \text{ lbf-ft-s}^2$ ). In the azimuth axis, four motors share the load instead of two as in the elevation axis. Assuming that each motor operates below 12.5% of maximum torque, the nonlinear spring constant ranges between 0.468 N-m/rad ( $0.345 \text{ ft-lbf/rad}$ ) and 3.23 N-m/rad ( $2.38 \text{ ft-lbf/rad}$ ). This corresponds to a natural frequency between 1.12 and 2.76 Hz. The small anti-resonance at 2.0 Hz is centered between these calculated frequencies. The anti-resonance is not as significant, however, as in the elevation axis. This may be due to cross coupling and structural damping between axes. An additional test using double the biased voltage and double the sinusoidal voltage amplitude was made to further investigate the anti-resonance reported above. Figure 18 presents the results of this test. The anti-resonance is pronounced during this test and validates prior observations.

## D. Unbiased Input Test — Azimuth Rate Loop

Three additional tests were performed on the azimuth axis using an unbiased sinusoidal input voltage with the elevation axis at angles 7, 45, and 88 degrees. The empirical Bode plots shown in Figs. 19, 20, and 21 indicate dramatic gain attenuation and phase lag at low frequencies and that the frequency response is not a function of elevation angle. Gain attenuation and phase lag are attributed to friction, backlash, and drive-associated nonlinearities. The contributions of each type of nonlinearity to the frequency response have not been determined. The frequency responses in Figs. 19, 20, and 21 are very different from the theoretical prediction shown in Fig. 4.

## E. Step Responses

Step response tests were performed on both elevation and azimuth rate loops. Normalized tachometer voltage is presented as a function of time in Figs. 22 and 23. The rise time and settling time of the elevation rate loop, 0.12 seconds and 0.37 seconds, respectively, are more than 20% longer than the predicted values shown in Fig. 5. The measured rate loop response gave a 1% overshoot compared to the 4% predicted overshoot.

The azimuth rate loop step response had similar characteristics to the elevation rate loop: rise time = 0.13 seconds; settling time = 0.34 seconds; percent overshoot = 0%. Again, the measured rise time and settling time are more than 20% longer than the predicted values in Fig. 6. Both the low percent overshoot and fast rise time exhibited by the rate loops are desirable performance characteristics.

## V. Conclusions

The frequency response of elevation and azimuth rate loops had several anomalies. Anti-resonances existed in both elevation and azimuth rate loops. These resonances were deduced to be fundamental torsional modes of vibration between the antenna and motor inertias. The calculated elevation axis torsional natural frequency between 1.87 Hz and 2.84 Hz was in good agreement with the observed resonance at 2.3 Hz. The calculated azimuth axis natural frequency between 1.12 Hz and 2.76 Hz was also in good agreement with the observed resonance at 2.0 Hz.

System bandwidth of elevation and azimuth rate loops, based on a biased input voltage, was 8.8 Hz and 12 Hz, respectively. These were similar to the predicted bandwidths of 9.1

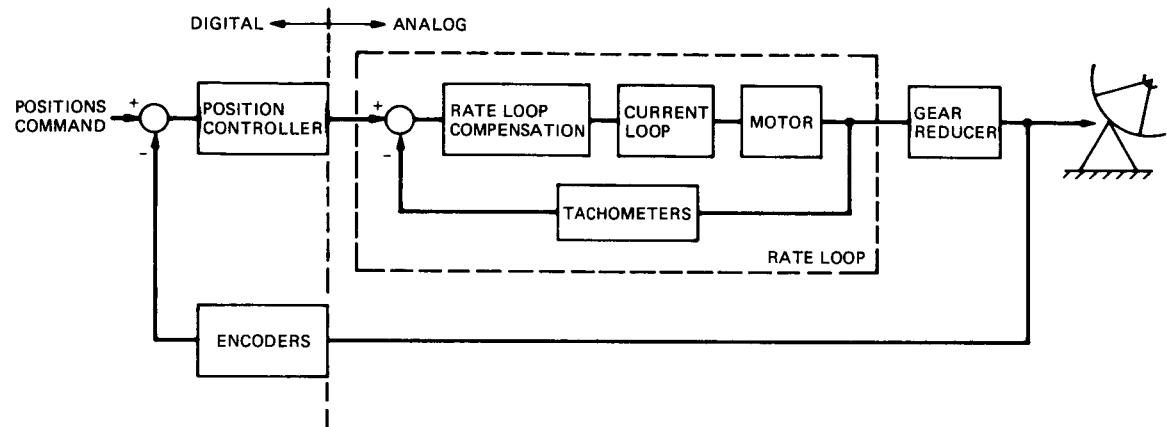
Hz and 9.4 Hz. Using an unbiased sinusoid voltage seriously degraded system bandwidth performance in the elevation axis by almost 90%. Unbiased voltage tests in the azimuth axis indicated that large nonlinearities exist such as friction, backlash, and other drive-associated nonlinearities.

Step response characterization tests indicate elevation and azimuth rate loops have fast rise times (less than 0.13 seconds), low or no overshoot, and settling times less than 0.4 seconds. These performance characteristics indicate that the rate loops respond well to command voltages.

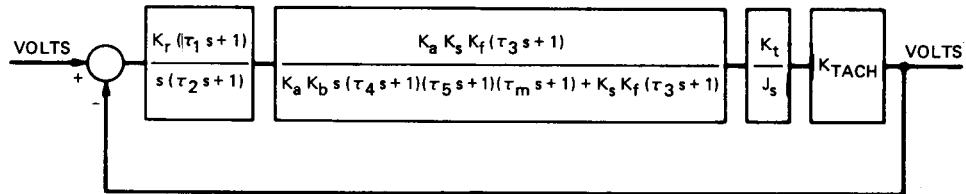
Both frequency and step response characterization indicate the linear mathematical model of the rate loop does not accurately describe system dynamics.

## References

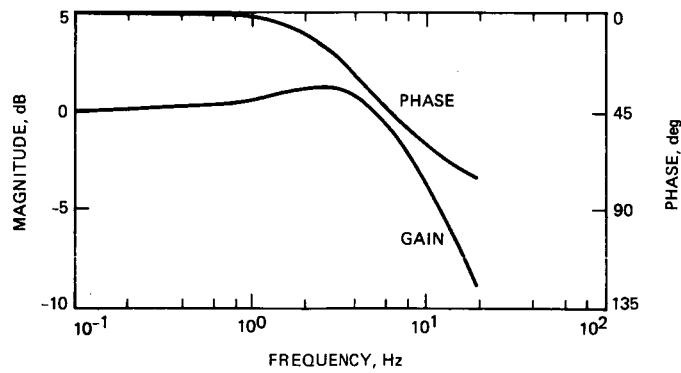
1. Dorf, R. C., *Modern Control Systems*, 3rd ed., Addison-Wesley, Reading, Mass., 1983.
2. Baumeister, T., Avallone, E. A., and Baumeister III, T., *Mark's Standard Handbook for Mechanical Engineers*, 8th ed., McGraw-Hill, New York, 1978.



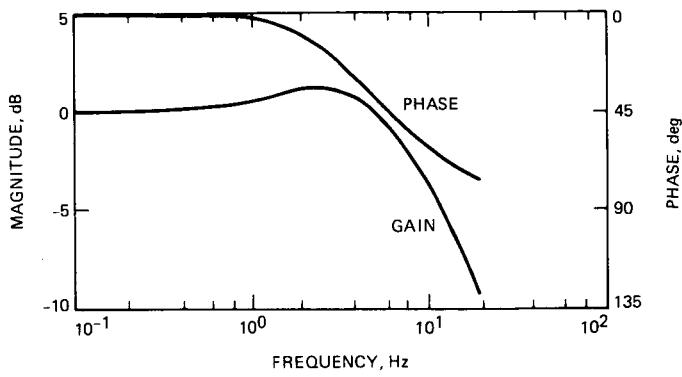
**Fig. 1. Position loop**



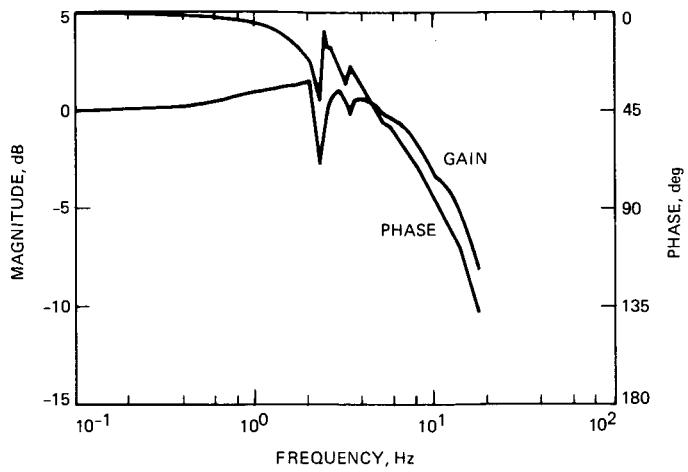
**Fig. 2. Rate loop**



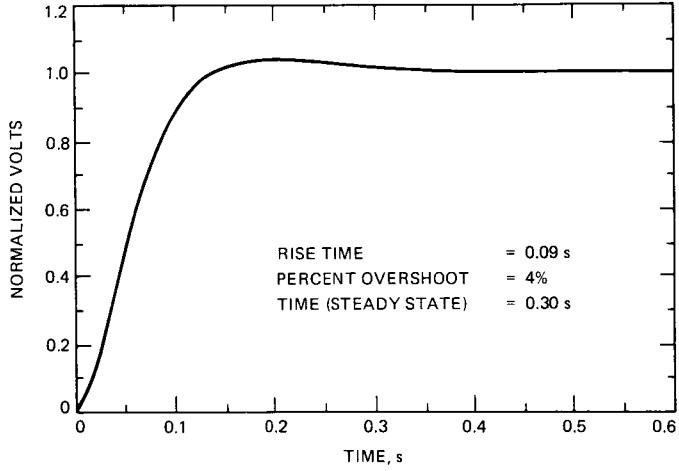
**Fig. 3. Simulated frequency response of the elevation rate loop model, DSS 15 antenna**



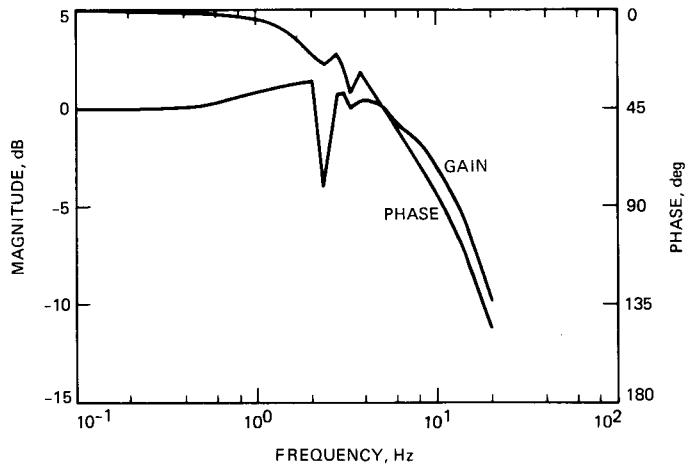
**Fig. 4. Simulated frequency response of the azimuth rate loop model, DSS 15 antenna**



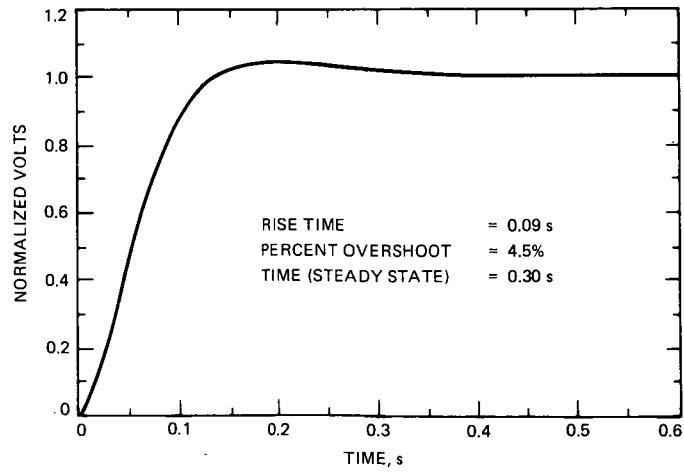
**Fig. 7. Biased input test — elevation axis loop, EL = 7 deg**



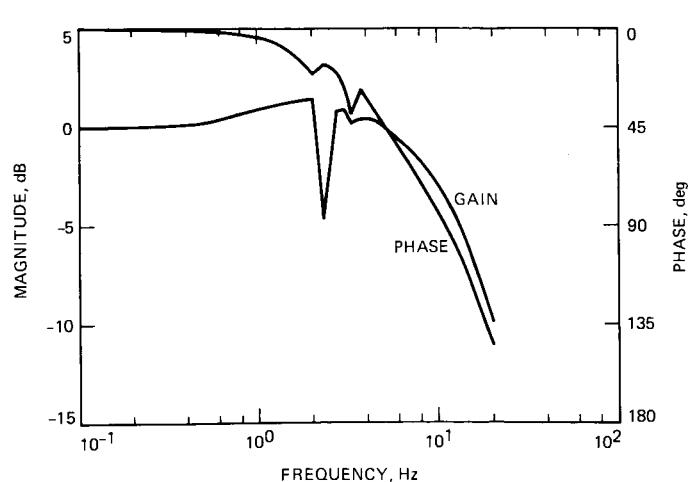
**Fig. 5. Simulated step response — elevation rate loop model, DSS 15 antenna**



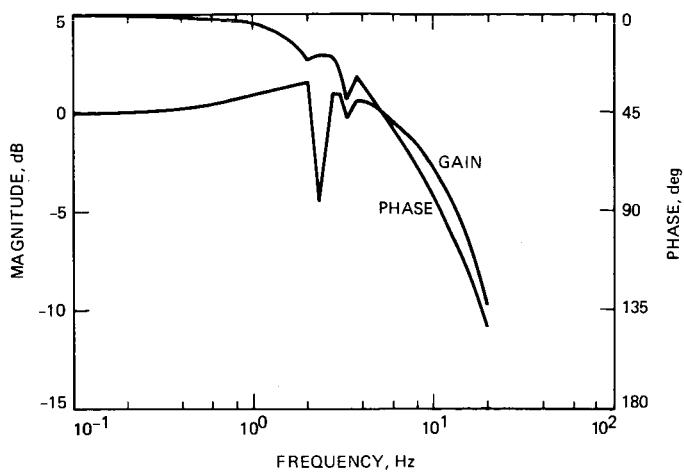
**Fig. 8. Biased input test — elevation axis loop, EL = 45 deg**



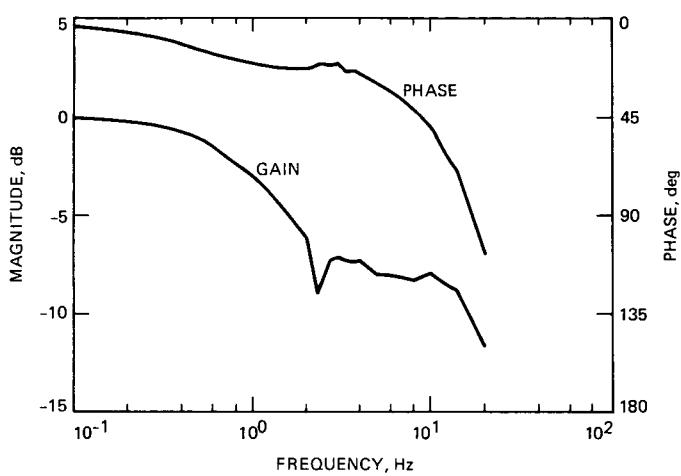
**Fig. 6. Simulated step response — azimuth rate loop model, DSS 15 antenna**



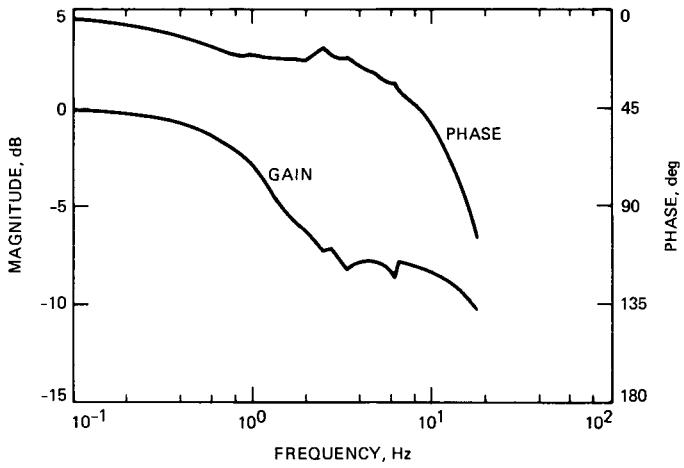
**Fig. 9. Biased input test — elevation axis loop, EL = 70 deg**



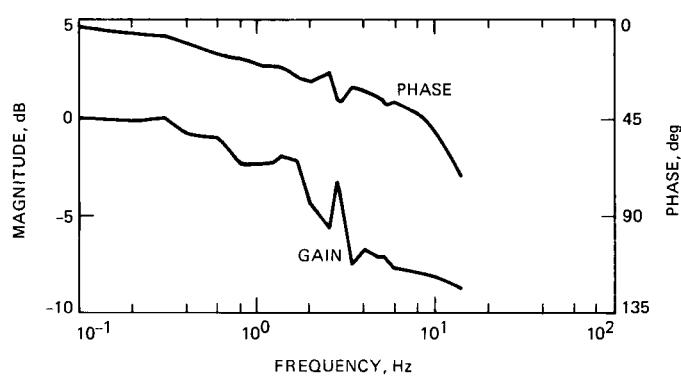
**Fig. 10. Biased input test — elevation axis loop, EL = 88 deg**



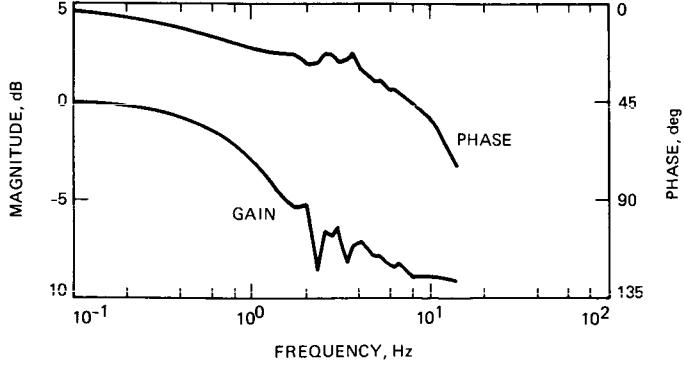
**Fig. 13. Unbiased input test — elevation axis loop, EL = 70 deg**



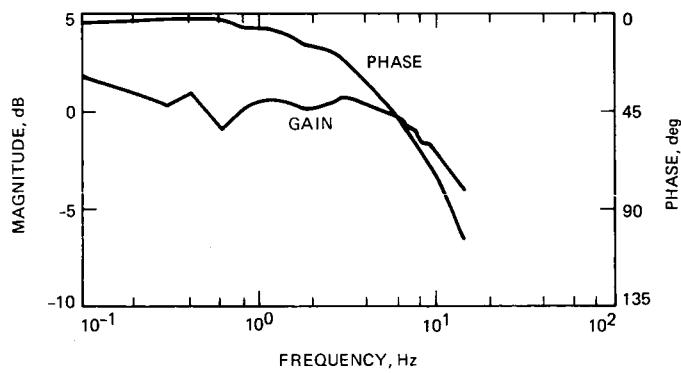
**Fig. 11. Unbiased input test — elevation axis loop, EL = 7 deg**



**Fig. 14. Unbiased input test — elevation axis loop, EL = 88 deg**

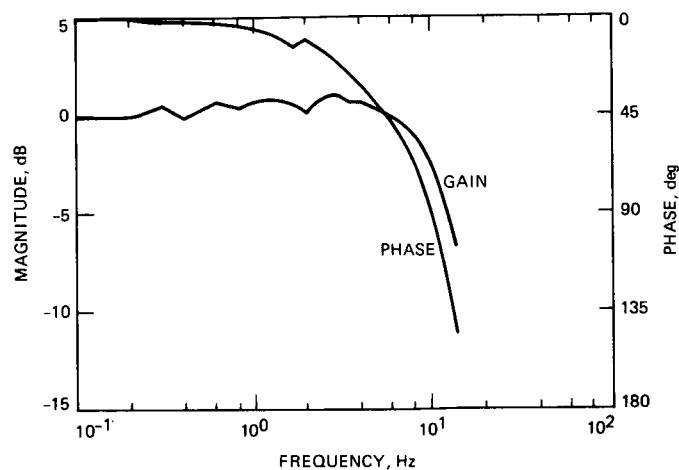


**Fig. 12. Unbiased input test — elevation axis loop, EL = 45 deg**

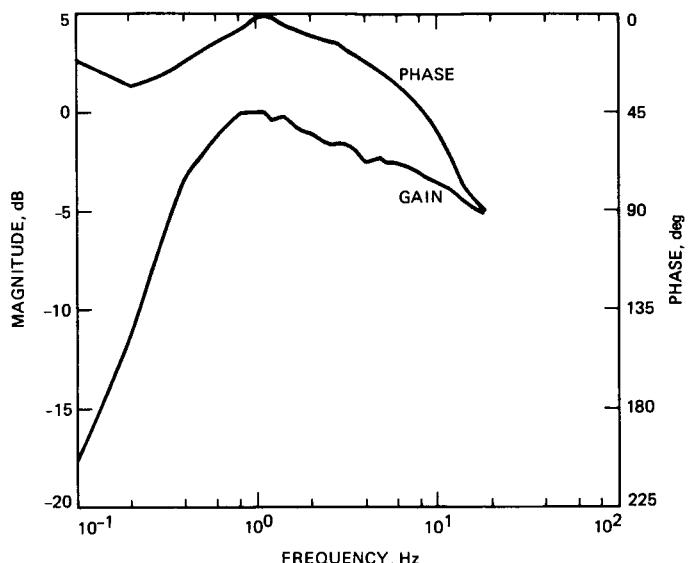


**Fig. 15. Biased input test — azimuth axis loop, EL = 7 deg**

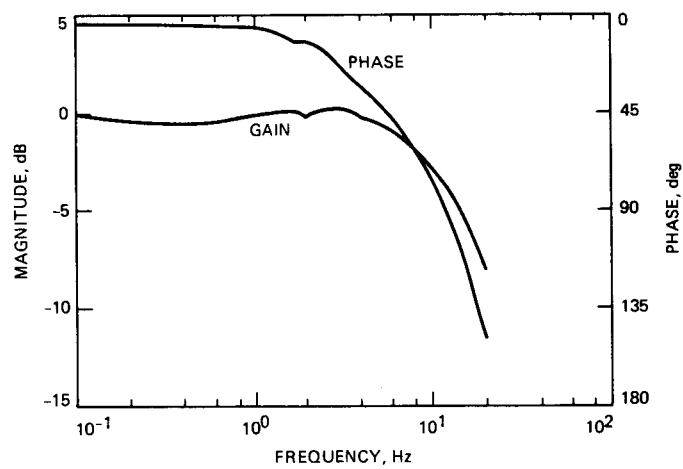
C - 3



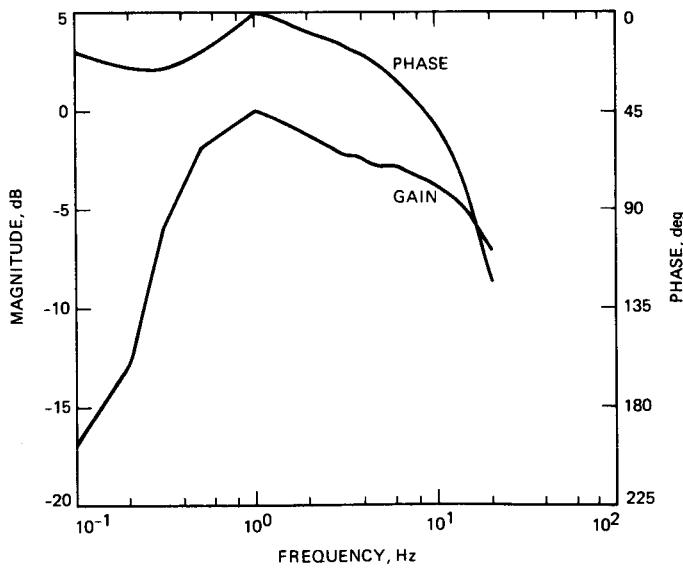
**Fig. 16. Biased input test — azimuth axis loop, EL = 45 deg**



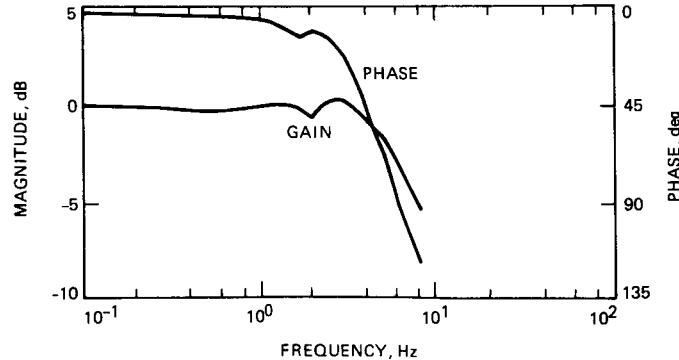
**Fig. 19. Unbiased input test — azimuth axis loop, EL = 7 deg**



**Fig. 17. Biased input test — azimuth axis loop, EL = 88 deg**



**Fig. 20. Unbiased input test — azimuth axis loop, EL = 45 deg**



**Fig. 18. Double input and double bias test — azimuth axis loop, EL = 45 deg**

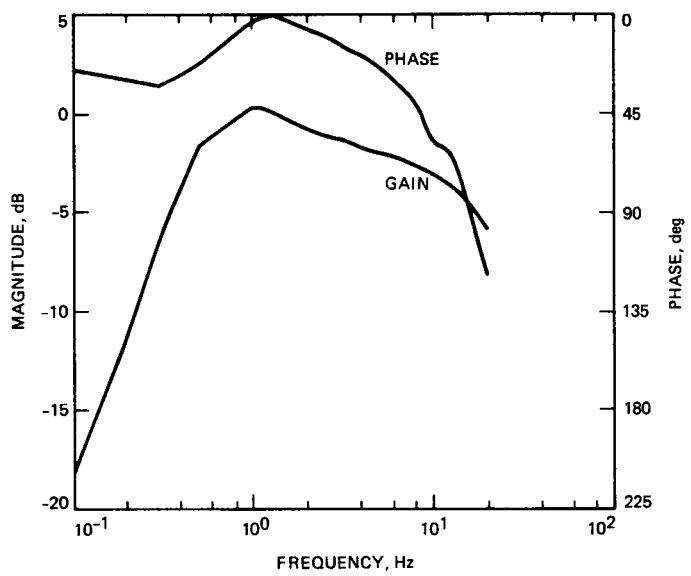


Fig. 21. Unbiased input test — azimuth axis loop, EL = 88 deg

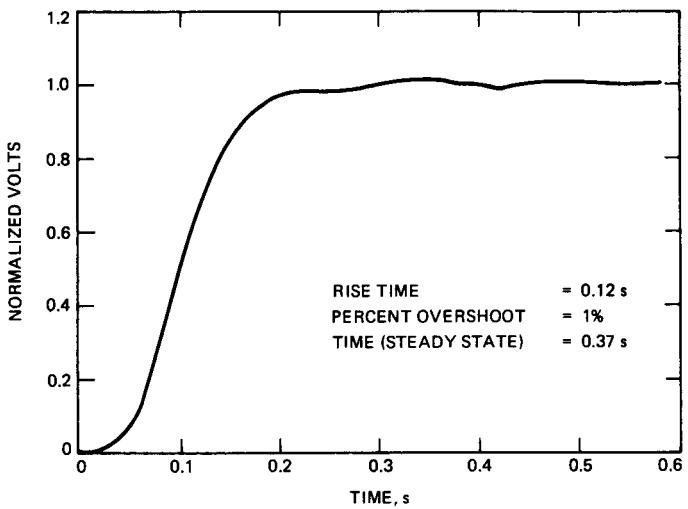


Fig. 22. Actual step response — elevation axis loop, DSS 15

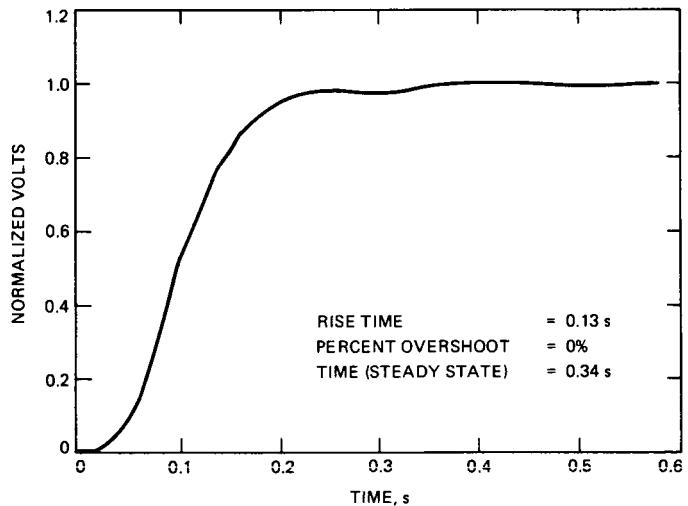


Fig. 23. Actual step response — azimuth axis loop, DSS 15

## Software Development for the VLA-GDSCC Telemetry Array Project

H. W. Cooper

Radio Frequency and Microwave Subsystems Section

L. R. Hileman

Telos Corp.

*Software for the VLA-GDSCC Telemetry Array (VGTA) Project is being developed in a new manner. Within the Radio Frequency and Microwave Subsystems Section, most microprocessor software has been developed using Intel hardware and software development systems. The VGTA software, however, is being developed using IBM PCs running consumer-oriented software. Utility software and procedures have been generated which allow the software developed on the IBM PCs to be transferred and run on a multibus 8086 computer.*

### I. Introduction

Intel was the only supplier of development hardware and software when the Radio Frequency and Microwave Subsystems Section first began using the Intel 8080 microprocessor for the control of electronic equipment. Over the years, several groups within the section have continued investing in and upgrading their Intel software development products.

While the Intel development system did serve its purpose, it was not a general purpose system in that both the hardware and the software were unique to Intel. New employees needed training on the system, and the system could not be used beyond the development of software because the Intel development system hardware could not be used as a target system.

### II. Software Development for Parkes-CDSCC Telemetry Array (PCTA)

The PCTA Project (Ref. 1), forerunner to the VGTA Project, used integral microprocessor computers to perform the monitor and control function of all of its electronic assemblies. In addition, an "array controller" performed as a message switcher by interfacing two CRT terminals to the various assemblies. While the assemblies all ran ROM-based firmware, the array controller contained floppy disk drives which ran the CPM-86 operating system that enabled the array controller to read and execute command sequences from the floppy disks.

When the PCTA project was started, the implementing group did not have the required Intel development equipment.

As an alternative, the PCTA software, written in Pascal, was developed in the "array controller" computer which ran the CPM-86 operating system. This procedure totally eliminated the need for Intel development products. While this method worked, it had difficulties. The Pascal MT+86 compiler errors were numerous, and much time was spent finding work-arounds for them. In addition, claims for easy ROM programmability using Pascal MT+86 were unfounded, and assembly language and 8087 support was poor. The system was floppy-disk based, resulting in slow compilation times. One application program exceeded 64K in size, which required a complicated work-around in order to run a ROM-based system.

### **III. VLA-GDSCC Telemetry Array (VGTA) Method**

What was needed was a hardware/software development method which was cost effective, well supported, free of errors, and efficient. After considering various options, the IBM PC system was selected since it used the same microprocessor, and because of the large availability of off the shelf software.

Given the development hardware, Microsoft "C," a high-level programming language, was selected for the following reasons:

- (1) Microsoft wrote the operating system for the IBM PC and would be expected to be capable of writing a "C" compiler for it.
- (2) "C" is a high-level structured language presently very popular in industry and in universities. The language includes a large standard library which performs most needed functions including port I/O, string functions, and memory management. The only problem initially noted with Microsoft "C" was its lack of real-time interrupt support, which has been solved by the development of JPL Software routines.
- (3) Microsoft "C" supports code and data memory requirements larger than 64K, and also provides 8087 numeric data processor support.
- (4) Microsoft also supplies a compatible macro assembler and symbolic debugger, both compatible with their "C" compiler.
- (5) The Microsoft products are well supported and are available off the shelf.

Once the development method was chosen, the main problem was how to make software designed to run on an IBM PC run on an 8086 multibus system. Microsoft "C" didn't make any claim to ROM programmability of their

code, and Microsoft "C" runs under the PC DOS operating system. The problems to be solved included how to transfer the code from the IBM PC to the 8086 computer, how to debug the software, and how to handle real-time interrupts. These problems have been solved through the use of utility software written at JPL. The utility software is used to transfer the compiled application program from the IBM PC environment to the target system. Transfer methods include floppy disks, PROMs, or downloading to RAM.

The "C" language (including Microsoft "C") itself does not provide interrupt handling. Even though all "C" functions are re-entrant, they end with a normal return, not an interrupt return. This problem has been solved by using simple assembly language routines which vector the interrupt, to the desired "C" routine, and then perform the required end-of-interrupt operations.

The VGTA retains the same monitor and control philosophy as did the PCTA. The ability to read and execute command sequences from the floppy disks was lost with the removal of the CPM-86 operating system from the array controllers. Therefore, it was decided to use the IBM PC as a smart terminal as well as a software development tool. Once again, selecting Microsoft "C" proved beneficial because it allows the IBM PC and Target Machine software to be written using the same language.

### **IV. Development Process and Required Equipment**

Software development can be described by the following steps:

- (1) Design the application program.
- (2) Code the program.
- (3) Compile the program.
- (4) Test the program.

In the preceding steps, all but testing can be done on any IBM PC. Initial testing of programs, if no real-time input/output is performed, can also be done to some degree on the IBM PC. But in general, testing must be done on the target computer. It is for this reason that the utility programs were written.

The following hardware and software were required:

- (1) IBM PC or equivalent with PC DOS.
- (2) Microsoft "C" and Microsoft macro assembler.

- (3) Target system consisting of:
  - (a) Intel 86/14 computer.
  - (b) 1/2 Mbyte RAM.
  - (c) PROM module.
  - (d) BLC 8222 floppy disk controller module.<sup>1</sup>
  - (e) 5-1/4 inch floppy disk drive.<sup>1</sup>
  - (f) Set of COLDBOOT, SIMDOS, and 957B PROMs.
- (4) Prolog PROM programmer.<sup>2</sup>
- (5) JPL utility software.

## V. Transferring IBM PC-Compiled Software

When a program is compiled to run on the IBM PC, a relocatable file, called an EXE file, is generated by the compile and link process. When the program is executed, PC DOS loads the EXE file into memory and then executes it. All input and output to the program is through PC DOS system calls, which are implemented via software interrupt 21H. The utility software performs the functions of locating the EXE file to operate at a specific memory location, loading the file into the target computer's memory, and simulating the interrupt 21H PC DOS system calls. The transferring of the program to the target computer's memory can be accomplished via floppy disk, PROMs, or downloading.

## VI. Memory Map of Target Computer

The following is a memory map showing where the various utility programs will be loaded:

- FFFF:F Top of memory.
- FFFF:0 Jump instruction to the start of COLDBOOT (F800:0).
- FC00:0 Start of Intel 957B monitor program (ROM).
- FD00:0 ROM image of SIMDOS. Moved to RAM by COLDBOOT during initialization sequence.
- F800:0 Start of COLDBOOT (ROM).
- 8000:0 Start of ROM area containing application program memory image.

Note: The area above 8000:0 is ROM (or unused); the area below is RAM.

- |        |   |
|--------|---|
| 1000:0 | Start of RAM area where application program is loaded from floppy disk, transferred from ROMs at 8000:0, or downloaded from IBM PC. |
| 07C0:0 | Start of RAM resident portion of SIMDOS. Transferred here during the boot-up sequence.  |
| 0780:0 | Start of the load area for the LOADER utility.  |
| 0200:0 | Start of RS-232 communication input buffer.   |
| 0000:0 | Start of interrupt vector table, bottom of memory.  |

## VII. Utility Software

All utility programs were written in "C," or assembly language, with the exception of the Intel 957B program. All programs are small and are easily modified.

**INTEL957B.** This program is an INTEL monitor program. It is designed to run on the 86/14 computer and to provide basic debugging services. This program is supplied in an unmodified form.

**EXE2ABS.** This program is written in "C," and runs on the IBM PC. It takes as input a relocatable EXE file generated by the Microsoft compiler/assembler/linker. This program converts the relocatable file into a memory image file designed to run at location 1000:0, although the load address may be changed if desired. The resultant absolute file can be loaded into the target computer via floppy disk using the LOADER utility, or it may be converted to a hex file for PROM programming using the ABS2HEX, and HEX2ROM utilities, or it may be downloaded to the target computer using the IBM2SBC utility. The EXE2ABS places a header at the beginning of the actual program which contains items such as the program length, program start address, a valid file ID mark, a file checksum, and code to initialize the various segment registers, stack pointer, and instruction pointer.

**SIMDOS.** This program, written in assembly language, runs on the target computer. It is designed to simulate PC DOS functions which the application program calls via software interrupt 21H. A ROM image of SIMDOS is moved into the RAM area during the initialization sequence of the COLDBOOT utility. Once moved, the SIMDOS initialization routine is called, which sets up the interrupt 21H vector.

Since application programs reside in ROM-based assemblies having no disk drives, the majority of the PC DOS functions relating to file and memory management are not needed.

<sup>1</sup>Required if floppy disk loading is desired. While the utility software is presently written for the National BLC 8222, the software can be modified for other controllers.

<sup>2</sup>Required to program PROMs. While the utility software is presently written for the Prolog programmer, it can be modified to control other programmers.

Only the functions required to support the RS-232 interface and a few miscellaneous others are implemented at this time. These include the following:

- 0 – Program terminate
- 1 – Keyboard input
- 2 – Display output
- 6 – Direct console I/O
- 7 – Direct console input without echo
- 8 – Console input without echo
- 9 – Print string
- B – Check standard input status
- 25 – Set interrupt vector
- 30 – Set DOS version number
- 35 – Get interrupt vector
- 40 – Write to a file or device
- 44 – I/O control for devices

For a detailed explanation of these functions, consult the PC DOS documentation.

*COLDBOOT*. This program, written in assembly language, is the first program to run on the target computer following a power on or reset. Upon start-up, COLDBOOT sets the baud rate, and initializes buffers required for the RS-232 I/O channel under interrupt control. Secondly, this program tests RAM from 0000:0 to 8000:0. If during the RAM test time (about 5 s) a key is pressed, a menu will appear following the completion of the RAM test, giving various options to the user. If a key has not been pressed, as is the case in normal operation, the program checks to see if a floppy disk is present and ready to be loaded. If so, the program on the floppy disk is loaded and executed. If no floppy disk is present, COLDBOOT checks to see if a ROM program is present. If a ROM program is present, it is transferred from the ROM area to the RAM area and then executed. If neither floppy disk nor ROM is present, the menu is presented. COLDBOOT always checks for a valid program ID mark, and performs a checksum of the RAM program prior to executing it. The menu items include the following:

- (1) Load from a floppy disk.
- (2) Transfer a ROM program to the RAM area.
- (3) Test RAM.
- (4) Execute the 957B monitor.

- (5) Perform a checksum of the RAM program.
- (6) Execute the RAM program.
- (7) Download a program into RAM.

*LOADER*. This program, written in assembly language, runs on the target computer. It is used to load a program into the target computer via a floppy disk. Using the IBM PC, the floppy disk must first be formatted without the /S option, thereby creating an empty disk. LOADER must be the first program written to the disk. Next, SIMDOS is written to the disk; and finally the application program, after being located to run at 1000:0 by the EXE2ABS program, is written to the disk. Following this procedure, the disk is ready to be loaded into the target computer. This process is not as complicated as it seems if a PC DOS batch command is used.

When the disk is inserted into the floppy drive of the target computer and the reset pressed, COLDBOOT will detect the presence of the floppy disk and load LOADER into address 0780:0, and then jump to 0780:0. The LOADER will then load the disk copy of SIMDOS into its proper place in memory, and then load the application program into location 1000:0. Valid program ID marks are checked, and a checksum of the application program is performed prior to program execution. By using the LOADER utility, a test version of either SIMDOS or the application program can be loaded into the memory of the target computer.

*ABS2HEX*. This utility program, written in "C," which runs on the IBM PC, converts an absolute file created by the EXE2ABS utility to a hex file which will then be used to program PROMs. ABS2HEX prompts the user for ROM size, and then generates hex files grouped as ROM images. The program also generates and records in the file the checksums for each ROM image.

*HEX2ROM*. This program, written in "C" and run on the IBM PC, reads a hex file generated by the ABS2HEX utility and then controls the PROM programming process. HEX2ROM is presently designed to control a Prolog PROM programmer but may be modified for others if desired. The user is directed by prompts during the programming process to perform checksums on the programmed ROMs.

## VIII. Conclusion

The software approach described above is presently being used to implement the VGTA Software. PCTA software, written in Pascal and compiled using CPM-86/Pascal MT+86, is presently being rewritten in "C" in order to eliminate bugs,

and to make the programs more manageable. Approximately 70% of the programs have been converted and tested. Results show a 20% decrease in code size, and at least a 50% increase in speed. Little assembly language programming has been

required in the application programs other than the interrupt linkage routines; this was not the case in the Pascal versions. So far, no errors have been detected with the Microsoft products.

## Reference

1. Brown, D. W., Cooper, H. W., Armstrong, J. W., and Kent, S. S., "Parkes-CDSCC Telemetry Array: Equipment Design," *TDA Progress Report 42-85*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 85-110.

D20-17

N87-15349

TDA Progress Report 42-87

July-September 1986

41519

## Real-Time Combiner Loss

M. K. Simon and A. Mileant  
Telecommunications Systems Section

*Telemetry signals from several channels are aligned in time and combined by the Real-Time Combiner (RTC) in order to increase the strength of the total signal. In this article, the impact of the timing jitter in the RTC on the bit/symbol error rate is investigated. Equations are derived for the timing jitter loss associated with the coded and uncoded channels. Included are curves that depict the bit/symbol error rate vs.  $E_b/N_0$  and  $E_s/N_0$  for some typical telemetry conditions. The losses are typically below 0.1 dB.*

### I. Introduction

The Real-Time Combiner (RTC) is part of the Baseband Assembly (BBA). Its purpose is to time align baseband signals from different antennas in order to increase the strength of the total combined signal. Figure 1 represents a block diagram of the overall downlink telemetry process of signal combining, subcarrier demodulation, symbol synchronization and data recovery.

If there were no timing jitter in the RTC, the maximum signal-to-noise ratio at the output of the RTC would be the sum of the SNRs at its input. However, timing jitter will decrease the output SNR from its maximum value. This decrease in combined SNR will in turn increase the bit/symbol error rate of the recovered data. In Ref. 1 the degradation in the RTC's output SNR was estimated. In this analysis, the degradation in bit/symbol error rate performance will be estimated for the rate 1/2, constraint length 7 convolutional code and the uncoded data cases. Expressions are derived for the timing jitter loss.

### II. Impact of the Timing Jitter on the RTC's SNR

As shown in Fig. 2, the operation of the RTC can be visualized as a weighted sum of  $L$  streams of binary data. At the discrete time instant  $i$ , the RTC produces at its output

$$x_i = \sum_{j=1}^L \alpha_j (s_{ji} + n_{ji}) \quad (1)$$

where  $\alpha_j$  is the  $j$ th weighting factor,  $s_{ji}$  is the signal amplitude of the  $i$ th sample of the  $j$ th signal with  $E\{s_{ji}\} = \sqrt{S_j}$  and  $n_{ji}$  is the amplitude of the additive thermal noise in the  $j$ th channel. This noise is modeled as a zero-mean Gaussian process with variance  $\sigma_j^2 = N_0 B_j$ , where  $B_j$  is the noise bandwidth at the  $j$ th input to the RTC. The noise processes in each of the channels are statistically independent of each other.

The transmitted data symbols are recovered by adding up all the samples belonging to each individual symbol. Let  $N_s$  be

the number of Nyquist samples per symbol, then, as shown in Fig. 2, at the end of each symbol the RTC produces

$$y_k = \sum_{i=1}^{N_s} x_i \quad (2)$$

Assuming *perfect time alignment* of the combined signals, inserting Eq. (1) in Eq. (2) and using the above definitions, the first two moments of the  $k$ th symbol,  $y_k$ , will be

$$E\{y_k\} = N_s \sum_{j=1}^L \alpha_j \sqrt{S_j} \quad (3)$$

$$E\{(y_k)^2\} = N_s^2 \sum_{j=1}^L (\alpha_j \sqrt{S_j})^2 + N_s \sum_{j=1}^L \alpha_j^2 \sigma_j^2 \quad (4)$$

The first term in Eq. (4) represents the expected value of the signal energy in the  $k$ th symbol, and the second term, the corresponding noise variance.

Imperfect timing in the RTC loops decreases the effective energy of the detected symbol. Since the RTC's loop update time is much larger than the symbol time, the alignment error is constant during many symbols. Let  $\epsilon_j$  be the normalized alignment error defined as follows

$$\epsilon_j \doteq 4|\delta_j|/N_{sc} \quad (5)$$

where  $\delta_j$  is the alignment error of the  $j$ th channel relative to the master channel ( $j = 1$ ) and  $N_{sc}$  is the period of the subcarrier waveform (both  $\delta_j$  and  $N_{sc}$  are in units of Nyquist samples and  $\delta_j$  is a random variable with the probability density function (pdf) given by Eq. (75) of Ref. 1). Using that equation, the pdf of  $\epsilon_j$  will be

$$p(\epsilon_j) = [1/(2\pi\sigma_{\epsilon j}^2)^{1/2}] [\exp\{-(\epsilon_j - \mu_j)^2/(2\sigma_{\epsilon j}^2)\} + \exp\{-(\epsilon_j + \mu_j)^2/(2\sigma_{\epsilon j}^2)\}] U(\epsilon_j) \quad (6)$$

where  $\mu_j = E\{4\delta_j/N_{sc}\}$ ,  $\sigma_{\epsilon j}^2 = (4/N_{sc})^2 \sigma_{Nj}^2$  and  $U(\cdot)$  is the unit step function. The closed-loop variance of the timing jitter,  $\sigma_{Nj}^2$ , is given by Eq. (53) of Ref. 1. It can be shown that for data modulating a squarewave subcarrier, the variance of  $\epsilon_j$  will be

$$\sigma_{\epsilon j}^2 = B_L \{[\text{erf}(\eta_1) \text{erf}(\eta_j)]^{-2} - 1\}/(2B_1) \quad (7)$$

where  $B_L$  is the bandwidth of the RTC loop and  $\eta_j \doteq (S_j/2\sigma_j^2)^{1/2}$ .

With *imperfect timing* in the RTC loops, the first two moments of the  $k$ th symbol conditioned on  $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_L)$  will be

$$E\{y_k | \epsilon\} = N_s \sum_{j=1}^L \alpha_j \sqrt{S_j} (1 - \epsilon_j) \quad (8)$$

$$E\{(y_k)^2 | \epsilon\} = N_s^2 \sum_{j=1}^L [\alpha_j \sqrt{S_j} (1 - \epsilon_j)]^2 + N_s \sum_{j=1}^L \alpha_j^2 \sigma_j^2 \quad (9)$$

$$\doteq E_k + N_k \quad (10)$$

Note that  $\epsilon_1 = 0$  by definition.

It can be shown (see Ref. 1, Eq. [66]) that optimum signal combining is obtained when the weighting factors  $\alpha_j$  are related as follows.

$$\alpha_j \sigma_j^2 / \sqrt{S_j} = \alpha_1 \sigma_1^2 / \sqrt{S_1}, \quad j = 2, \dots, L \quad (11)$$

Since scaling all  $\alpha$ 's by a constant factor does not change the SNR at the output of the RTC, we set  $\alpha_1 = 1.0$  and readily obtain the optimum values for all the remaining weighting coefficients. Namely, we make

$$\alpha_j = \frac{\sqrt{S_j}}{\sqrt{S_1}} \frac{\sigma_1^2}{\sigma_j^2} = \gamma_j / \sqrt{m_j}, \quad j = 2, \dots, L \quad (12)$$

where  $m_j \doteq S_1/S_j$  and  $\gamma_j \doteq \sigma_1^2/\sigma_j^2$ .

With these preliminary steps completed, we are now ready to evaluate the conditional SNR at the RTC's output. Let  $R_k \doteq E_k/N_k$  be the SNR at the end of the  $k$ th symbol (see Eq. (10)). Then, using Eq. (9),  $R_k$  conditioned on  $\epsilon$  is

$$R_k(\epsilon) = \frac{N_s [(\sum \alpha_j \sqrt{S_j})^2 - 2(\sum \alpha_j \sqrt{S_j})(\sum \alpha_j \sqrt{S_j} \epsilon_j)]}{\sum \alpha_j^2 N_0 B_j} + \frac{N_s (\sum \alpha_j \sqrt{S_j} \epsilon_j)^2}{\sum \alpha_j^2 N_0 B_j} \quad (13)$$

where all summations are from  $j = 1$  to  $j = L$  and, again,  $\epsilon_1 = 0$  by definition.

Using the definitions of  $m_j$  and  $\gamma_j$ , the optimum values of  $\alpha_j$  (Eq. (12)) and simplifying, the conditional symbol SNR can be expressed in terms of  $S_1$  and  $B_1$ , namely

$$R_k(\epsilon) = \frac{N_s S_1 [(\Sigma \gamma_j/m_j)^2 - 2(\Sigma \gamma_j/m_j)(\Sigma \gamma_j \epsilon_j/m_j)]}{N_0 B_1 (\Sigma \gamma_j/m_j)} + \frac{N_s (\Sigma \gamma_j \epsilon_j/m_j)^2}{N_0 B_1 (\Sigma \gamma_j/m_j)} \quad (14)$$

$$= R_b [\beta - 2\Sigma \gamma_j \epsilon_j/m_j + (\Sigma \gamma_j \epsilon_j/m_j)^2/\beta] \quad (15)$$

where

$$R_b \doteq 2E_{s1}/N_0 = 2R_s \quad (16a)$$

and

$$\beta \doteq \sum_{j=1}^L \gamma_j/m_j \quad (16b)$$

Note that  $\beta > 1$  when  $L > 1$ . Equation (15) was obtained using the following relations:  $N_s = 2B_1/r$  and  $S_1/r = E_{s1}$  where  $E_{s1}$  is the symbol energy in the first channel and  $r$  is the symbol rate. Note that with no timing jitter in the RTC, the output SNR is

$$R_k = R_b \beta \quad \text{for all } k \quad (17)$$

So,  $\beta$  represents the factor by which the ideal combined SNR is larger than the SNR in channel 1.

### III. Evaluation of the Loss Due to Timing Jitter in the RTC

Timing jitter in the RTC loops has the effect of decreasing the effective SNR of the combined signal. This decrease of SNR in turn increases the bit/symbol error rate of the recovered data. In order to avoid confusion of terms, we define *degradation factor* as the factor by which the combined SNR is decreased due to timing jitter in the RTC, whereas we define *jitter loss* (like “radio loss”) as the factor by which the SNR has to be increased in order to achieve a specified bit/symbol error rate. The evaluation of the degradation factor (decrease in combined SNR) requires averaging over  $\epsilon$  of the conditional SNR in Eq. (15). This evaluation was done in Ref. 1. In this analysis, the jitter loss, namely, the degradation in bit error rate performance, is determined.

The evaluation of the jitter loss requires two steps. First, the particular coding scheme has to be specified. Then, averag-

ing on the bit/symbol error rate is performed over the timing jitter pdf, Eq. (6). In our analysis two telemetry cases will be considered: (a) the rate 1/2, constraint length 7 convolutionally coded and (b) uncoded data.

#### A. Convolutional Code

For the rate 1/2, constraint length 7 convolutional code, an expression for bit error rate has been empirically found and is given as follows:

$$P_b = f_b(E_b/N_0) = \begin{cases} C_1 \exp(C_2 E_b/N_0), & E_b/N_0 > 0.8989 \\ 1/2, & E_b/N_0 \leq 0.8989 \end{cases} \quad (18)$$

where  $C_1 = 85.7469$  and  $C_2 = -5.7230$ . The term  $E_b/N_0$  is the bit energy to noise spectral density ratio at the input to the Maximum Likelihood Decoder (MCD). Note that  $E_b/N_0 = 2E_s/N_0 = R_b$  of Eq. (15).

#### B. Uncoded Data

For uncoded data the symbol error rate is given by

$$P_s = f_s(E_s/N_0) = (1/2)\operatorname{erfc}[(E_s/N_0)^{1/2}] \quad (19)$$

where  $\operatorname{erfc}(x)$  is the complementary error function. Note that now  $E_s/N_0 = R_b/2$  of Eq. (15).

Since the data rates considered in our analysis are higher than the RTC’s bandwidth, or, in other words, the loop timing error process  $\delta_j$  is slow and essentially constant over a bit/symbol time interval, the average bit/symbol error rate is estimated using the so called “high-rate model,” namely

$$P_i = \int_0^1 \dots \int_0^1 f_i(R_k(\epsilon) \prod_{j=2}^L p(\epsilon_j) d\epsilon, i = \begin{cases} b, & \text{for coded channel} \\ s, & \text{for uncoded channel} \end{cases} \quad (20)$$

where  $R_k(\epsilon)$  is given by Eq. (15) and  $p(\epsilon_j)$  by Eq. (6).

For the coded channel, the average bit error rate as given above can be obtained in a closed form as follows. Assuming that the RTC has doppler compensation,  $\mu_j$  in the pdf of the normalized timing jitter (Eq. [6]) will be zero. This is our first simplification in the evaluation of Eq. (20). A second simplification results in assuming that the contribution of the rightmost term in the numerator of Eq. (15) is much smaller than

that of the middle term (i.e.,  $E\{\epsilon_j\} > E\{(\epsilon_j)^2\}$ ). By discarding the right-most term in the numerator of Eq. (15) we will obtain the upper bound of  $P_b$ . Inserting now these simplified Eqs. (6) and (15) with (18) in Eq. (20), integrating and simplifying, we finally obtain the average bit error rate at the output of the MCD, namely

$$P_b = C_1 \exp(C_2 R_b \beta) \prod_{j=2}^L I_j \quad (21)$$

where

$$I_j = \exp(\nu_j^2) \operatorname{erfc}(\nu_j) \quad (22)$$

$$\nu_j = \sqrt{2} C_2 R_b \gamma_j \sigma_{\epsilon_j} / m_j \quad (23)$$

and  $R_b$  is given by Eq. (16a).

For the uncoded case, the average degraded symbol error rate can be obtained by first approximating the complementary error function of Eq. (18) by  $\operatorname{erfc}(x) \approx \exp(-x^2)/\sqrt{\pi x}$ . Then we expand this function in a Taylor series about  $\epsilon_j = 0$  ( $j = 2, \dots, L$ ). Finally, we take the expected value of the series expansion and keep the first few terms. Performing the above steps we find after some algebra the desired  $P_s$ , which for  $L = 3$  becomes

$$P_s = (1/2) \operatorname{erfc}(\sqrt{R_s \beta}) [1 + \sqrt{8/\pi} (R_s + 1/\beta)(\lambda_2 + \lambda_3) + 2(R_s^2 + R_s/\beta + 3/4\beta^2)(\lambda_2^2 + \lambda_3^2 + \lambda_2 \lambda_3)] \quad (24)$$

where again  $R_s \doteq E_{s1}/N_0$  (Eq. [16a]) and  $\lambda_i \doteq \gamma_i \sigma_i / m_i$ . For  $L = 2$ , we simply set  $\lambda_3 = 0$  in the above equation.

When, in the region of specified (desired) bit/symbol error rate, both curves of  $P_i$  vs. SNR (the ideal and that degraded by the RTC's timing jitter) are parallel, the *timing jitter loss* can be obtained from the following expressions.

For the coded channel

$$L_b = 10 \log_{10} [R_b \beta / (E_b / N_0)_{\text{ref}}], \text{dB} \quad (25)$$

where  $(E_b / N_0)_{\text{ref}} = \ln(P_b / C_1) / C_2$ , ( $\ln(\cdot)$  denotes natural logarithm) and  $P_b$  is the degraded bit error rate obtained from Eq. (21).

For the uncoded channel

$$L_s = 10 \log_{10} [R_s \beta / (E_s / N_0)_{\text{ref}}], \text{dB} \quad (26)$$

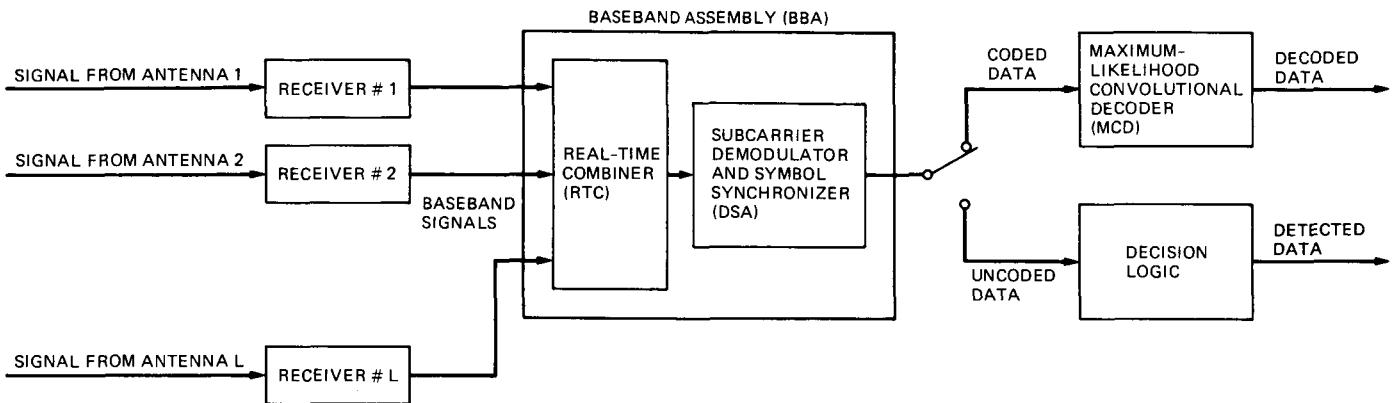
where  $(E_s / N_0)_{\text{ref}} = [\operatorname{erfc}^{-1}(2P_s)]^2$ ,  $\operatorname{erfc}^{-1}$  is the inverse complementary error function and  $P_s$  is the degraded symbol error rate obtained from Eq. (24).

## IV. Conclusion

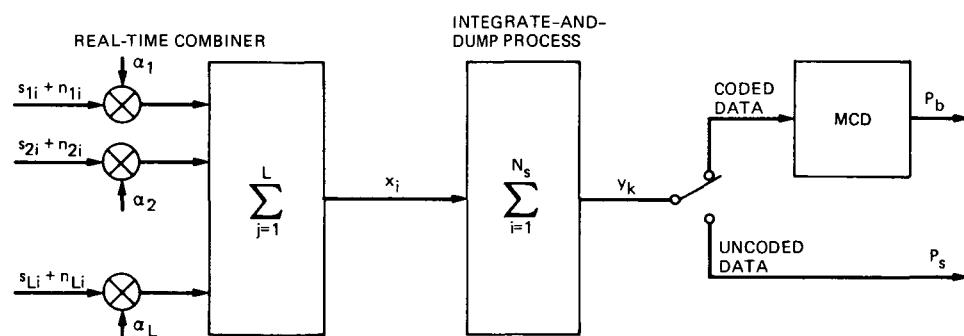
In the above analysis the impact of the timing jitter in the RTC on the bit/symbol error rate was evaluated. Equations were obtained for the degraded average bit/symbol error rate vs. SNR in the reference channel for the rate 1/2, constraint length 7 convolutionally coded and uncoded data. The result of the above analysis are illustrated in Figs. 3 through 6 for several typical telemetry conditions. The losses are typically below 0.1 dB.

## Reference

1. Simon, M. K., and Mileant, A., *Performance Analysis of the DSN Baseband Assembly (BBA) Real-Time Combiner (RTC)*, JPL Publication 84-94, Rev. 1, Jet Propulsion Laboratory, Pasadena, Calif., May 1, 1985.



**Fig. 1. Downlink process of signal combining, subcarrier demodulation, symbol synchronization and decoding of telemetry data**



**Fig. 2. Simplified representation of the signal combining and decoding process**

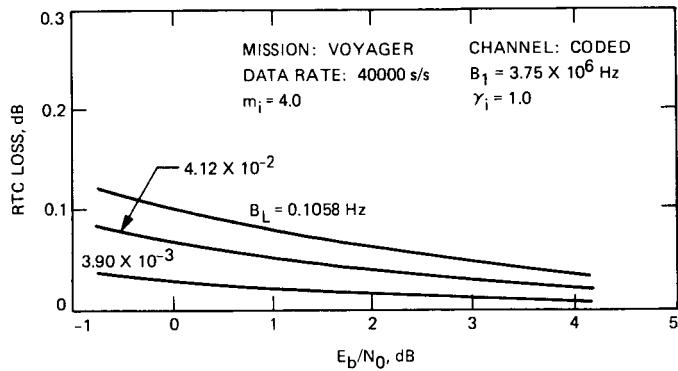


Fig. 3. The RTC loss vs master antenna  $E_b/N_0$ ,  $L = 2$

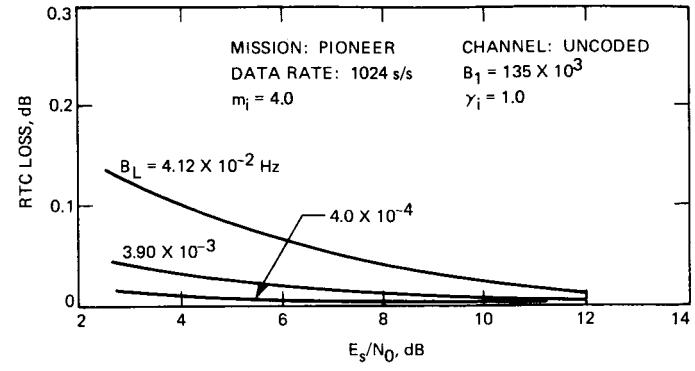


Fig. 5. The RTC loss vs master antenna  $E_s/N_0$ ,  $L = 2$

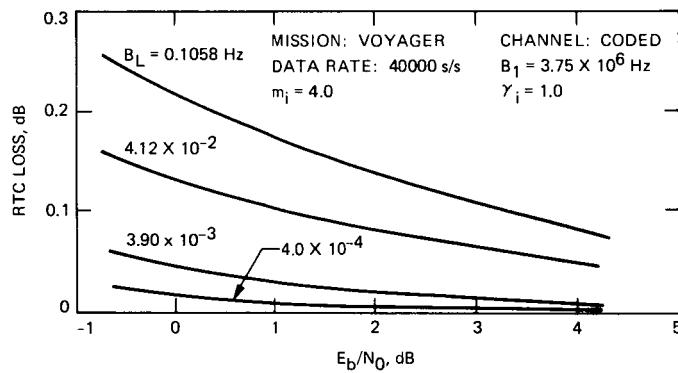


Fig. 4. The RTC loss vs master antenna  $E_b/N_0$ ,  $L = 4$

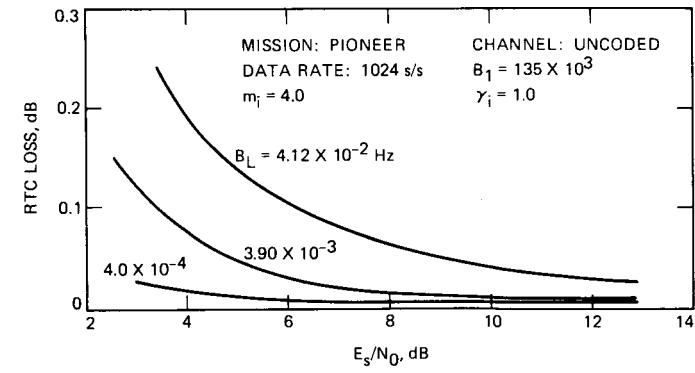


Fig. 6. The RTC loss vs master antenna  $E_s/N_0$ ,  $L = 3$

D21-32  
N87-15350 P8

# PPM/NAR 8.4-GHz Noise Temperature Statistics for DSN 64-Meter Antennas, 1982–1984

S. D. Slobin and E. M. Andres  
Radio Frequency and Microwave Subsystems Section

*From August 1982 through November 1984, X-band downlink (8.4-GHz) system noise temperature measurements were made on the DSN 64-m antennas during tracking periods. Statistics of these noise temperature values are needed by the DSN and by spacecraft mission planners to assess antenna, receiving, and telemetry system needs, present performance, and future performance. These measurements were made using the DSN Mark III precision power monitor noise-adding radiometers located at each station. It is found that for DSS 43 and DSS 63, at the 90% cumulative distribution level, equivalent zenith noise temperature values fall between those presented in the earlier (1977) and present (1983) versions of DSN/Flight Project design documents. Noise temperatures measured for DSS 14 (Goldstone) are higher than those given in existing design documents and this disagreement will be investigated as a diagnostic of possible PPM or receiving system performance problems.*

## I. Introduction

The precision power monitors (PPMs) installed on the three DSN 64-m antennas use a noise-adding radiometer (NAR) to monitor the system noise temperature, and a signal level estimator (SLE) to monitor the spacecraft signal level. Descriptions of these instruments will be published in the *TDA Progress Report* in the future. A description and analysis of NAR operation are given in Refs. 1 and 2. Over the two-year period from August 1982 through November 1984, over 240,000 system noise temperature measurements were made using the NARs during spacecraft tracking periods. While noise temperature variations are primarily due to weather effects, these variations may also be indicative of problems in the

microwave/receiver systems. Thus, the PPM can serve as a useful real-time diagnostic indicator of station performance.

During Mark IV implementation, the three 64-m DSN stations were not operational for the full period of August 1982 through November 1984. The valid data periods for the stations are as follows:

- DSS 14: September 1982–June 1983
- DSS 43: August 1982–September 1984
- DSS 63: August 1982–November 1984

Because of the shorter data collection period, DSS 14 statistics may not be directly comparable to those of DSS 43 and

DSS 63. In particular, the months of July and August (which are missing) are characterized by high absolute humidity and thunderstorm activity.

Typically, an NAR determines system operating noise temperature ( $T_{OP}$ ) by pulsing a noise diode on and off and injecting this square-wave noise into the main signal channel of the receiving system. In the DSN application described above, a typical noise diode modulation rate is 10 Hz and the injected noise is typically  $\sim 1$  K. Two fundamental equations govern NAR operation. They are, for  $T_{OP}$ :

$$T_{OP} = T_N / (Y - 1) \quad (1)$$

where

$T_{OP}$  = system operating noise temperature, K (defined at system input reference plane)

$T_N$  = noise diode injected noise temperature, K (defined at system input reference plane)

$$Y = (V_2 + \alpha V_2^2) / (V_1 + \alpha V_1^2), \text{ ratio}$$

$V_2$  = detector output voltage, noise diode on, V

$V_1$  = detector output voltage, noise diode off, V

$\alpha$  = detector nonlinearity constant,  $V^{-1}$  (= 0 in an ideal detector)

and for noise temperature resolution:

$$\Delta T_{OP} = 2T_{OP}(1 + T_{OP}/T_N)/\sqrt{\tau B} \quad (2)$$

where

$\tau$  = measurement time, s

$B$  = predetection bandwidth, Hz

For DSN operational use, the value of this injected signal level ( $T_N$ ) is kept low to avoid contamination of the very low system noise temperature ( $\sim 20$ – $25$  K). On the other hand, for high resolution (small  $\Delta T_{OP}$ ), a large value of  $T_N$  is required.

In operation, the PPM performs the following steps to determine  $T_{OP}$ :

- (1) The PPM switches the waveguide system to an ambient load ( $\sim 300$  K) and calibrates a large ( $\sim 50$  K) noise diode according to

$$T_N = T_{OP,\text{amb}}(Y - 1) \quad (3)$$

and

$$T_{OP,\text{amb}} = T_P + T_E \quad (4)$$

where

$T_E = T_M + T_F$  (equivalent noise temperature of the receiver, K)

$T_M$  = maser noise temperature, K

$T_F$  = followup receiver noise temperature, K

$T_P$  = physical temperature of ambient termination, K.

- (2) The PPM switches the waveguide out the horn and uses the calibrated 50 K noise diode to determine  $T_{OP}$  on the cold sky according to Eq. (1).
- (3) Knowing the sky  $T_{OP}$ , the PPM chooses a smaller operational noise diode to give a  $T_{OP}$  resolution ( $\Delta T_{OP}$ ) of 0.1 dB (2.33%) of the value determined in Step (2).

Thus, for example, if cold sky  $T_{OP}$  is determined to be 40 K, the required  $\Delta T_{OP}$  is 0.93 K, and the  $T_N$  chosen for operational use will have a value somewhat greater than 0.35 K for an integration time of 10 seconds and a bandwidth of  $10^7$  Hz. In this case, a 0.5 K noise diode would be chosen. This amount of injected noise will not seriously degrade receiving system performance.

Clearly, an inaccurate calibration of the large noise diode when looking at the ambient load will result in errors in determining  $T_{OP}$ . A possible source of calibration error is due to receiver nonlinearities (saturation) at large  $T_{OP}$  ( $>300$  K). In this case, the measured  $Y$  factor will be smaller than it should be, and the value of  $T_N$  determined from Eq. (3) will thus be too low. This will ultimately result in a lower than actual determination of  $T_{OP}$  according to Eq. (1).

## II. Data Reduction

Tapes of PPM data were generated monthly for each 64-m station. Only the X-band data were analyzed, as S-band weather effects are minimal and contribute insignificantly to increased system noise temperature. The data contained numerous PPM diagnostics and antenna tracking parameters, including date, time, spacecraft identification, maser assembly number, receiver assembly number, system noise temperature, standard deviation of  $T_{OP}$ , and antenna pointing (local hour-angle and declination). For each data point, antenna azimuth and elevation values were calculated based on the station's

latitude and the local hour-angle and declination of the spacecraft. Each data point (typically one every 40 seconds) was examined for validity according to the following criteria:

- (1)  $T_{OP}$  between 10 K and 300 K.
- (2) Nonzero local hour-angle and declination.
- (3) A changing (nonstuck) hour-angle.
- (4) Elevation greater than 3 deg.
- (5) Spacecraft declination in the range of +60 deg to -60 deg.
- (6) Standard deviation of  $T_{OP}$  between 0 and 2 K.

Short of invoking overly sophisticated logic or hand-picking the data points, it was felt that the six criteria above would sort out the vast majority of bad data points. Of the 282,547 total points recorded, 37,715 (or 13.35%) were discarded as failing one or more of the above criteria. Table 1 shows the percentage of points taken on each spacecraft and the percentage of points taken in each 10-deg elevation-angle interval.

Initial inspection of the histogram of noise temperature values showed a small cluster of points between 10 and 14 K for DSS 14, and an "abnormal looking" distribution between 17 and 20 K for DSS 63. In each of these cases, the data points were discarded. They amounted to 1.5% and 3% of the totals, respectively. No adjustment of this sort was made for the DSS 43 data points.

An attempt was made to create a zenith noise temperature model from the elevation-angle (nonzenith) points in the raw data. A  $1/\sin(\text{elevation})$  model was used. The method was as follows:

- (1) The lowest  $T_{OP}$  value in each station's valid data was assumed to be the best weather and highest elevation angle point of the test period.
- (2) From the value above, a zenith ground noise contribution (3 K) and zenith clear sky atmosphere contribution were subtracted. The remainder is the constant antenna noise contribution due to the horn, waveguide, maser, and cosmic background contribution (considered to be constant over a wide range of conditions).
- (3) An equivalent zenith system noise temperature ( $T_{OP,90}$ ) is calculated according to the equation

$$T_{OP,90} = T_{OP,\theta} (\sin \theta) - \Delta T_{\text{ground}} - (T_{\text{const}} + T_{\text{ground},\theta}) (\sin \theta - 1) \quad (5)$$

where

$$\begin{aligned} T_{OP,\theta} &= T_{OP} \text{ at elevation angle } \theta \text{ (raw data)} \\ \Delta T_{\text{ground}} &= \text{decrease in ground contribution} \\ &\quad \text{when moving to zenith} \\ &= T_{\text{ground},\theta} - 3.0 \\ T_{\text{const}} &= \text{constant antenna contribution deter-} \\ &\quad \text{mined in Step (2)} \\ T_{\text{ground},\theta} &= 3.0 + 5.0 [(90 - \theta)/90] \end{aligned}$$

This technique attempts to create a uniform condition (zenith-looking) by which the three 64-m antennas can be compared.

### III. Results

Figure 1 shows the probability density function (histogram) of the X-band system noise temperature for the three 64-m stations. The integral of the PDF, over all noise temperatures, is by definition 1.0. In the data set presented here, all data points below 20 K have been removed as described above.

Note that DSS 43 shows a narrower range of system noise temperatures than do the other two stations. This does not necessarily indicate that "better weather" existed at DSS 43. Because of the preponderance of Voyager-related points (and Jupiter, Saturn, and Uranus have had significant southerly declinations for the past several years), DSS 43 would have tracked generally at higher elevation angles than the other two sites. This high elevation bias would result in lower atmospheric noise temperature values, even during clear sky conditions.

Figure 2 shows the cumulative distribution (CD) of system noise temperature (the integral of the histograms in Fig. 1) at the three 64-m sites. These CDs are for all valid data and represent a mix of elevation angles, spacecraft, and tracking periods. If one were to blindly infer weather statistics from these curves, it would appear that DSS 63 has the best weather, and DSS 14 the worst. At the 90% "confidence level" (CD = 0.90), the system noise temperatures are 44.5 K (DSS 14), 41.4 K (DSS 43), and 38.5 K (DSS 63). It must be reiterated that the three stations do not have similar distributions of elevation angle (see Table 1), and that data were not collected for a long enough time period to generate truly long-term statistics.

Figure 3 shows the equivalent zenith CDs for the distributions in Fig. 2. The method of generating these points is described in Section II. The equivalent zenith points for Fig. 3 were generated one by one; the distributions in Fig. 2 were not

modified as a whole to create the zenith distributions. Also shown in Fig. 3 are system noise temperature CDs given in both the present (1983) and earlier (1977, "old 810-5") versions of DSN/Flight Project interface design documents<sup>1</sup>. The 810-5 curves are presented for comparison only; it is premature to attempt a replacement of the 810-5 models with the relatively short-term PPM measurements. The statistics of the older interface design documents are still being used in the Voyager telemetry prediction and performance (TPAP) computer program. It has been noted that the TPAP program, using the "old 810-5" model, generally predicts worse telemetry performance than is actually observed. Indeed, at the 90% confidence level (a commonly used performance reference point), the  $T_{OP}$  values at both DSS 43 and DSS 63 were predicted (old 810-5) to be higher than those measured by the PPM and higher than those predicted in the present version of the DSN/Flight Project design documents. At DSS 14, however, the measured PPM values appear to be substantially higher than both design-document models. If this is indeed an erroneous result, one possible explanation is that the short data gathering period (10 months) has somehow biased the data with worse than average weather conditions. Also, the method of creating the equivalent zenith model may be at fault, but then the zenith statistics at all stations would be

<sup>1</sup>Deep Space Network/Flight Project Interface Design Handbook, TDA document 810-5, Rev. D, TCI-40, Rev. B, pp. 1-31 (internal document), Jet Propulsion Laboratory, Pasadena, Calif., 1977 and 1983.

wrong in the same way, not just those at Goldstone. Other possible sources of DSS 14 errors are receiver system linearity, unstable noise diode outputs, defective PPM operation, or systematic, incorrect operational calibration of the PPM. Further data collection and analysis using the Mark IV PPMs will help to resolve this question.

#### IV. Conclusions

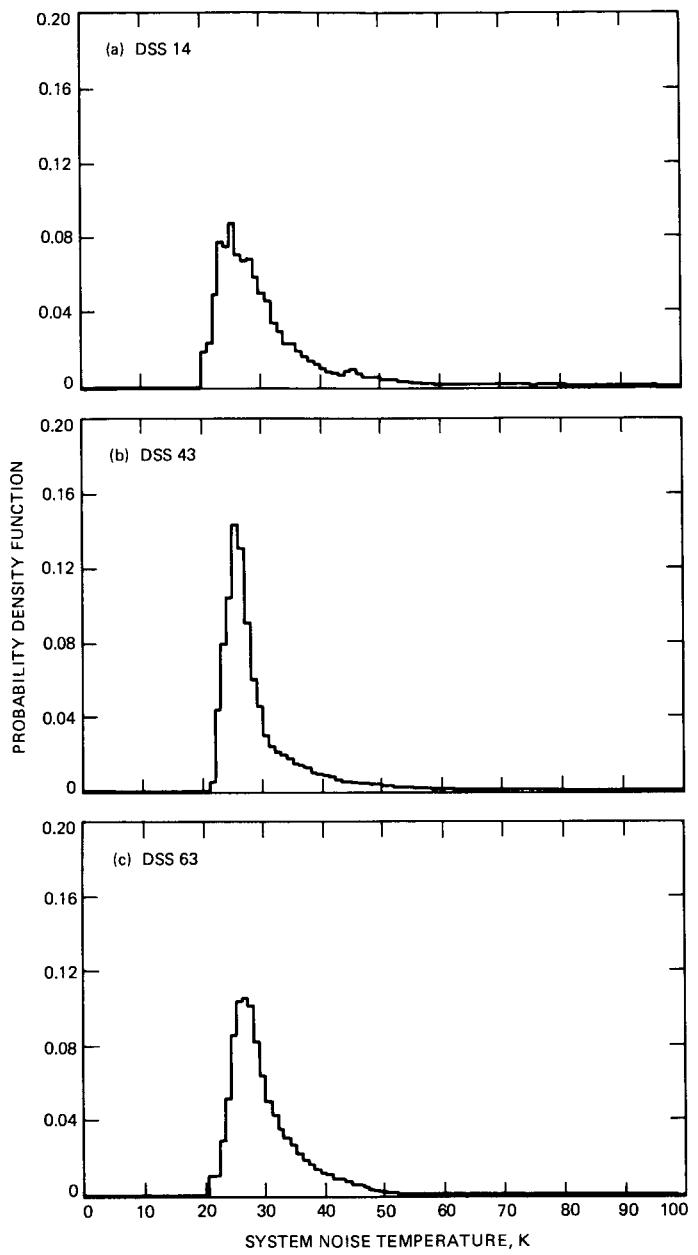
Several years of system noise temperature data were collected using the noise-adding radiometers in the PPMs at the three DSN 64-m stations. At the 90% confidence level, it is seen that the equivalent zenith  $T_{OP}$  measurements made by the PPM fall between the earlier (1977) DSN/Flight Project Interface Design Document model (used for Voyager) and the more "optimistic" model developed in 1983. The zenith results presented for Goldstone (DSS 14) show substantially higher  $T_{OP}$  values than either the 1977 or 1983 models. The cause of this is presently unknown, but may be related to the short period of data recording. Qualitative observations of weather at the three 64-m sites indicate that Goldstone has the most benign weather conditions and should experience the lowest noise temperature increases. The resulting statistics, then, are puzzling, and may indicate problems with the PPM itself, the receiving system, or with certain operational procedures. These problems will be fully investigated in order to guarantee reliable and accurate data output from these instruments in the future.

#### References

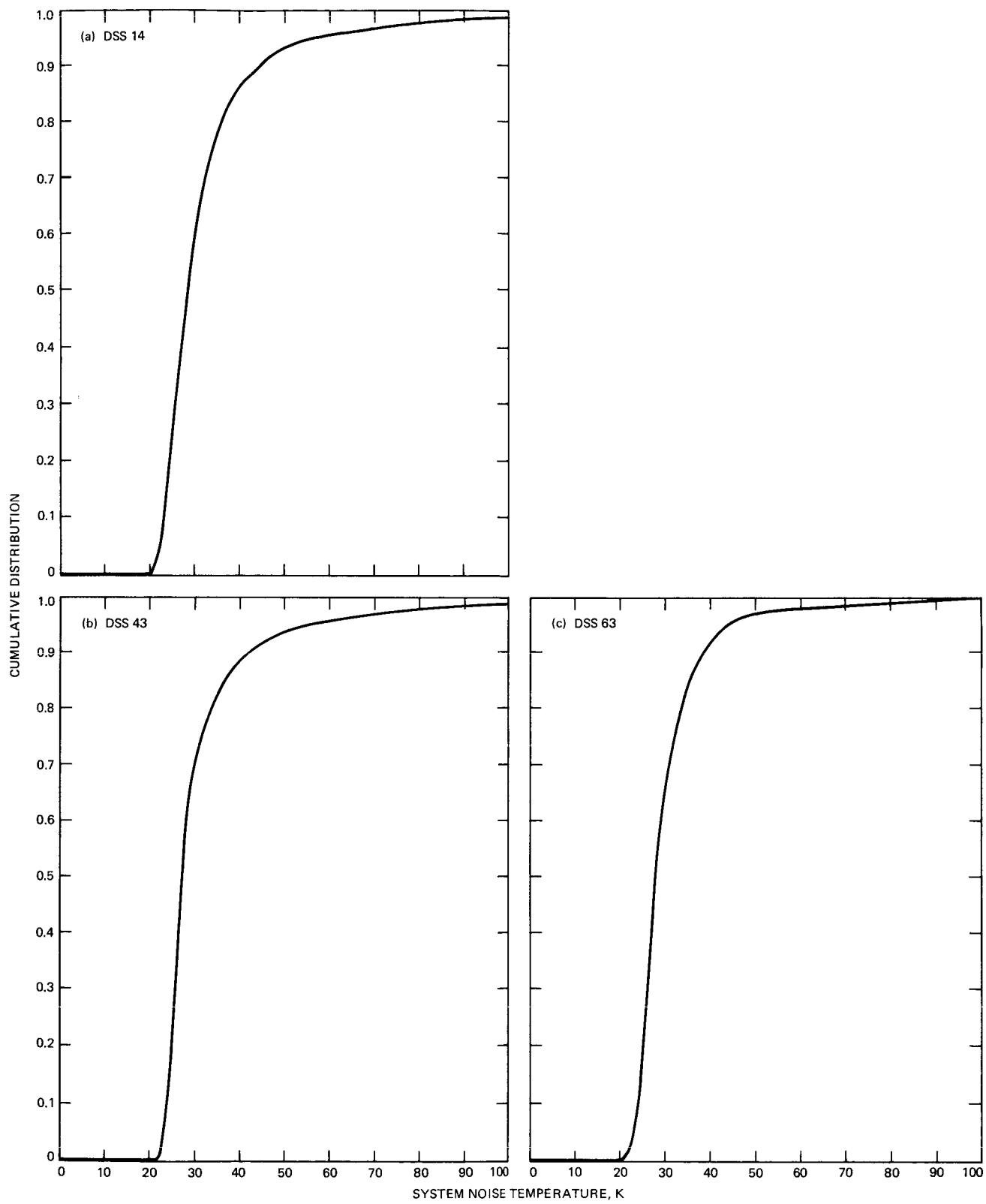
1. Batelaan, P. D., Goldstein, R. M., and Stelzried, C. T., "A Noise Adding Radiometer for Use in the DSN," *Space Programs Summary 37-65, Vol. II*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 66-69.
2. Stelzried, C., "Noise Adding Radiometer Performance Analysis," *TDA Progress Report 42-59*, Jet Propulsion Laboratory, Pasadena, Calif., October 15, 1980, pp. 98-106.

**Table 1. Fractional occurrence of spacecraft tracks and antenna elevation angle**

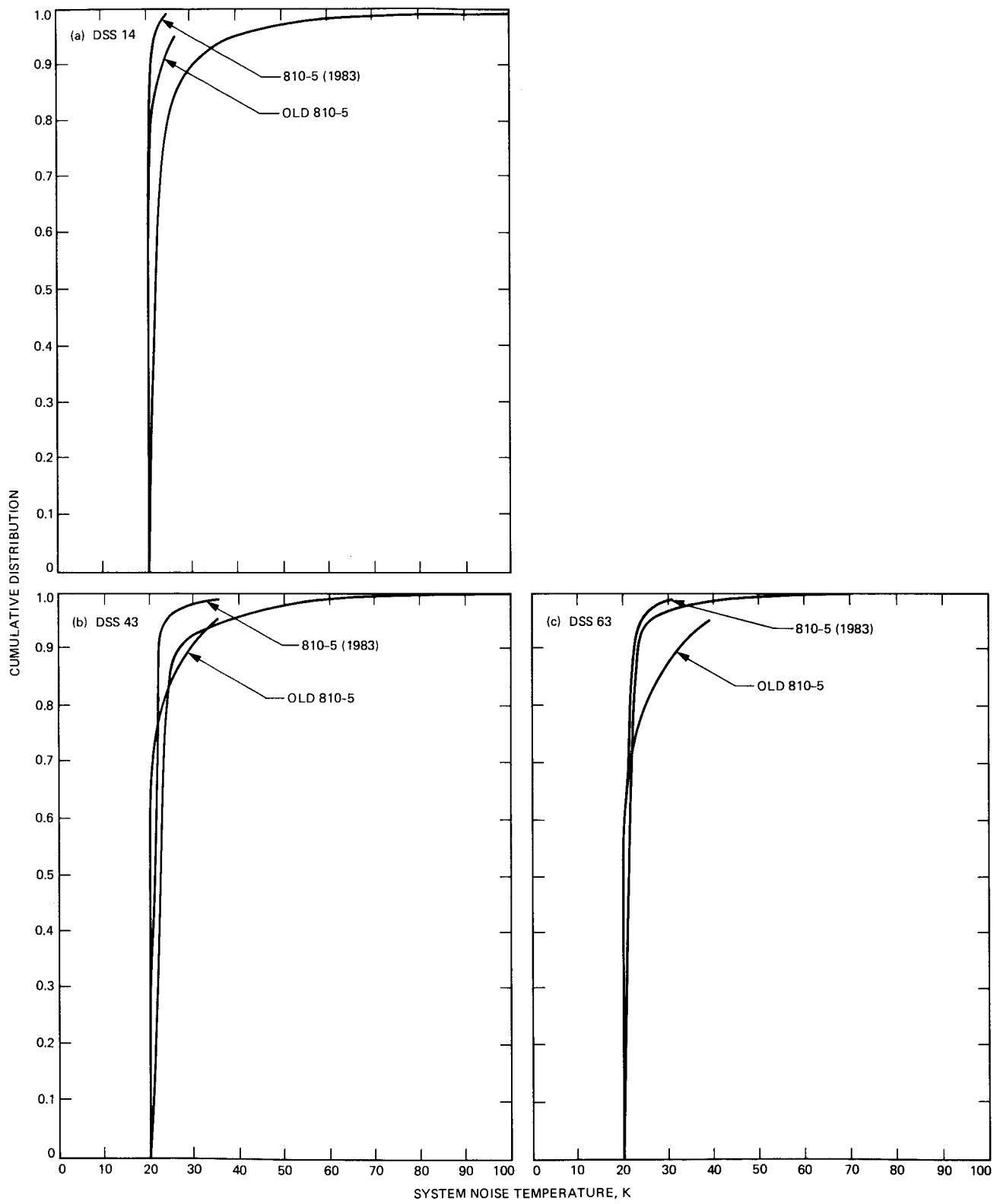
Parameter	DSS 14	DSS 43	DSS 63	All Stations
<b>Spacecraft</b>				
12 (Pioneer 12)	0.0881	0.1307	0.0390	0.0773
23 (Pioneer 10)	0.0036	0.0034	0.0031	0.0033
24 (Pioneer 11)	0.0026	0.0001	0.0000	0.0005
26 (Viking Lander 1)	0.0001	0.0000	0.0000	0.00004
31 (Voyager 1)	0.4040	0.3163	0.4952	0.4213
32 (Voyager 2)	0.5015	0.5492	0.4627	0.4973
90 (Helios 1)	0.0001	0.0003	0.0000	0.0001
<b>Elevation Angle, deg</b>				
0 – 10	0.0336	0.0158	0.0392	0.0308
10 – 20	0.1467	0.1084	0.1940	0.1580
20 – 30	0.2098	0.1361	0.3023	0.2322
30 – 40	0.3249	0.1700	0.1866	0.2082
40 – 50	0.1011	0.2221	0.1012	0.1391
50 – 60	0.1124	0.1163	0.1667	0.1403
60 – 70	0.0705	0.1445	0.0096	0.0638
70 – 80	0.0009	0.0867	0.0005	0.0276
80 – 90	0.0000	0.0000	0.0000	0.0000
<b>Point Count</b>				
Total	49568	98433	134546	282547
Good	47516 (95.86%)	76820 (78.04%)	120496 (89.56%)	244832 (86.65%)
Bad	2052 (4.14%)	21613 (21.96%)	14050 (10.44%)	37715 (13.35%)



**Fig. 1. X-band system noise temperature, probability density function (histogram)**



**Fig. 2. X-band system noise temperature, cumulative distribution, all elevation angles**



**Fig. 3. Equivalent zenith X-band system noise temperature, cumulative distribution, with 810-5 models for comparison**

# Riveted Panel Surface Measurement Using Photogrammetry

W. D. Merrick, V. B. Lobb,

F. L. Lansing and F. W. Stoller

Ground Antenna and Facilities Engineering Section

*Two riveted antenna panels on rings number 3 and 9 were removed from the 34m antenna at DSS-15, fixed in the leveled position and the surface was photographed indoors. The results from this pilot photogrammetric demonstration and diagnostics of panel surface contours, are presented. The photogrammetric network for each panel incorporated eight photographs, two from each of four camera stations and observed over 200 targets. The accuracy ( $1\sigma$ ) of the XYZ coordinates for the error ellipsoids was  $\pm 0.013$  mm (0.0005 inch). This level of precision relative to the object size corresponds roughly to 1 part in 250,000 which is superior to conventional dial sweep-arm template techniques by at least a factor of 4.*

## I. Introduction

During the construction of the NASA-JPL 34m High Efficiency (HE) antennas at DSS-15 and DSS-45, two identical sets of surface panel inspection jigs using sweep-arm templates were provided. One set was for use during panel fabrication and the other set for field inspection at the construction site prior to panel installation on the antenna. Single panel samples from each of the nine antenna rings differed in their measured surface accuracies when inter-compared in both sets of sweep-arm jigs.

Photogrammetric measurements (PGM) as a pilot demonstration on two sample panels were performed to resolve the above jig measurement disparities. By use of a more precise metrology technique such as PGM the demonstration provided also the means for validation of the cost, accuracy, speed of measurement, and the work-time spans required for future projects.

Geodetic Services Inc. (GSI), under contract with the Jet Propulsion Laboratory, performed the PGM field measurements, and data processing. The contractor, GSI, was selected by means of a wide survey of U.S. and Canadian vendors. GSI has introduced and practiced many new technologies in close-range photogrammetry over the past two decades. These include: the bundle method of processing photogrammetric triangulation, calibration procedures, instrumentation upgrades and field practices. As a result, the contractor has developed the best available precision and versatile specialized photogrammetric equipment, (Refs. 1-5).

This article describes the measurement procedure, the hardware and the software technologies.

## II. Test Procedure

Similar to installation upon the operational antenna structure, the two shaped-contour panels were tested while sup-

ported upon the adjusting 4-6 corner screws as shown in Figs. 1-4. The small panel (from antenna ring No. 3) has the dimensions of 2.0 X 2.7 m (6.5 X 7 ft) and the larger panel (from outer ring No. 9) has the dimensions of 2.1 X 2.7 m (7 X 9 ft). Small stick-on photogrammetric retrotargets were installed on the faces of the two panels at locations close to those previously measured by the jigs. The positions of reference targets were measured to provide a tie to global antenna coordinates.

Four points on each panel were leveled and the targets near the other adjustment screws were brought to the proper design heights. Then, the Z coordinates (vertical elevations) of all targets were observed with the same precision level (a calibrated Swiss-made Wild N3 instrument). The test conditions were indoors, in shaded and air conditioned space. The panels were supported by steel fixtures that rested upon a thick concrete slab well isolated from room vibrations.

The four key photogrammetry elements needed are: (1) the camera, (2) the film digitizer (comparator), (3) the computer, and (4) the software. The procedure may be summarized as follows: A two-person crew, used a special photogrammetric camera to take pictures of the several hundred small, self-adhesive 1-cm diameter retrotargets installed upon each panel. The completion of these photos required only a few hours at the demonstration site. Note that actual photogrammetric accuracy improves in proportion to the square root of the number of photos.

The coordinates of each target image were "read" automatically on a special film digitizer at the contractor's facility in Melbourne, Florida. After reading the film data, the (X, Y, Z) coordinates of each target are computed. The rays from all camera stations converging at a given target determine an individual error ellipsoid for that target. Using least squares techniques, the semi-axes of the ellipsoids are obtained for each target. The PGM technique proved advantageous as a high accuracy, low cost with short measurement time technique (less than one hour per camera network). The demonstration employed the main eight ingredients needed for the highest accuracy in photogrammetry which are listed below:

- (1) The use of the bundle method in triangulation with self-calibration of all elements of each independent photogrammetric network.
- (2) The use of highly convergent photography for maximum geometric strength.
- (3) The use of a long focal length, large format camera.
- (4) The use of a large number of exposures in simultaneous bundle data reduction.

- (5) The use of roll film and vacuum registry against a stable camera platen equipped with integral reseau projectors.
- (6) The use of analytical compensation for the variation of image distortion with object distance (e.g., diffraction).
- (7) The use of a computer based Simulator for planning of optimum PGM networks and operational fieldwork.
- (8) The calibration of the lens' radial, decentering and other distortions, as well as the camera cone and platen imperfections.

### III. Test Equipment

The single GSI CRC-1 camera with a 240 mm (9.45 in.) focal length lens and cone was moved between the camera stations. The selected film was Kodak Tech Pan rated at approximately 200 ASA. By utilization of the proper developer, a single step grey scale was attained which suppressed the extraneous details of the object being measured. Even if the object had been in bright sunlight, the full target details would have been retained. A strobe flash located proximate to the camera lens, was used to 'turn on' the adhesive-backed retrotargets (fabricated from 3M type 7610 beaded material).

Convergent imaging geometry was employed. Before conducting the field tests, a computer based Simulator was used to optimize and identify the positions of the desired camera stations as in Fig. 5, to enhance the geometric strength and reliability of the triangulation. Photography of the object from movable stations was simulated to provide the maximum possible convergence angles; those which improve the accuracy in the Z direction so that it approaches the accuracy of the X, Y coordinates.

The camera was located upon an air-actuated vertical service lift at approximately 3.8 m (12.5 ft) above the panel corners. With the tripod, camera and operator in place upon the platform, the outriggers were raised or lowered and the service lift was moved manually by two men.

Absolute scale was established by means of known distances between targets attached to a thermally stable surveying tape positioned across the face of each panel as shown in Fig. 2. The film was developed on site under controlled processing temperature.

The images of the targets were automatically read and the digitized coordinates recorded on the GSI AutoSet-1 mono-comparator located at the contractor's facility. Since the comparator is mounted upon a thick granite slab, it is the

only non-mobile (but relocatable) element in the PGM system. This fully automated film reader has a resolution of under 0.1 micron (4 micro inch) and is accurate to about 0.4 micron (16 micro inch) in each of the two coordinate axes. The Simulator results in the images of the center dots of the targets being about 75 microns in diameter. The area of each dot image is digitized into 400 to 750 pixels. The statistical centroid is computed, and the image is automatically re-centered by the comparator before the coordinates are recorded, thereby eliminating the reading bias error associated with a human operator.

Multiple camera roll angles were used to improve the accuracy of each network. When properly handled in photogrammetric computations, rotating the camera about the axis of the lens serves to cancel out all lens, lens cone and camera error except platen non-flatness (which is measured by other techniques and is fully compensated for in the reduction of the data).

The deleterious effects of distortion of the film emulsion and film base, are eliminated by reading, within the camera, the position of 25 "target-like" calibration images that are projected through the film backing onto every film frame by patented reseau projectors (mounted in the precision film platen made from temperature-stable material). The reseau images are located so that the target images are never more than 35 mm from the nearest reseau image. Four additional calibration images are projected onto the front face of each film frame from projectors located near the camera lens.

The outer ring of each of the retrotargets validates that the film was in contact (actuated by vacuum) with the special stable camera platen at the time of film exposure. If the outer ring of the image of the target shows as a crescent instead of a full circle, the suspect image/point is read and recorded with a weight of zero. At a later time, by measuring the dimensions of the ring image, an operator completes the computations required to recover the precise data for the center of that imperfect image.

The comparator-measured X, Y coordinates were then processed through a series of computer programs which provide rigorous simultaneous, least squares spatial triangulation and bundle adjustment of all measured data. "Bundling" denotes the reduction of large matrices into a plethora of smaller sparse ones that may be handled by simple known methods.

The self-calibration processes described above act to minimize the volumes of the error ellipsoid(s) which result from the non-intersection of the convergent rays that determine the Cartesian coordinates of each and every camera station and

target/point. The three values for the semi-axes of each error ellipsoid are computed. Ultimately, the X, Y, Z space coordinates for each target/point (and each camera lens station) are determined. The accuracy of the coordinates is independent of the nature of the photographed object.

#### IV. Test Time Span

For future planning purposes, it was essential to establish work-time relationships for estimating PGM measurement time spans, workhours and operational costs. Table 1 delineates the activities that occurred during the test period.

Experience shows that repetition of a given photogrammetric task usually requires less time to complete than the prototype operation. For example, our "first time" film reading (23 cm X 23 cm format) took 4 frames per hour. A second time iteration with the same crew would take 8 frames per 1.5 hour. The triangulation and bundling software takes about 0.5 hour to complete in either case. Our "first time" task required field photography and film development by two men working 10 hours on site, plus travel time.

In summary, the PGM demonstration took 6 hours to measure two panels, 5 days to complete the whole task, and 1 minute/frame to film the last network (includes camera movement).

#### V. System Simulation

Figure 5 illustrates the relative camera-panel locations for simulation and Figs. 6-9 give the Simulator output for optimum PGM accuracy. Note that the targets are shown covering the full area of the panels, but the operator has opted for fewer targets for easier screen perusal. The Simulator data appear at the left hand side of Figs. 6-9; and it further generates expected standard deviation (1-sigma) accuracy, image diameters, depth of field, etc. The accuracy that would be expected from networks where one film frame was exposed at each of four camera stations is given in Table 2 as 0.033 mm (0.0013 inch) for the Z-axis.

The initial formulations of the Simulator were conservative. When implemented, the resultant field accuracy exceeded the requirements. References to Tables 2 and 3 shows that the ratio of the accuracy predicted by the Simulator to the delivered accuracy, was 0.033 mm/0.013 mm or 2.6/1. This improved ratio results because doubling the number of films exposed at each station increased the accuracy ratio by a factor of 1.41.

In Building G-84 at the Mars Deep Space Station, heavy equipment stored along the walls prevented the service lift

(and supported camera) from occupying the optimum camera stations as determined by the Simulator. Since this was also a "first time" demonstration, the crew exposed two photographs at each camera station with differing flash lamp intensities (100 Ws and 200 Ws). Two camera roll angles of 135 and 235 degrees were used also to cancel out potential error contributions from the camera.

The Z axis runs near the center (or lowest point) of each panel, perpendicular to the plane taken through three points (located near the corners of each panel). In Fig. 5 these corner reference points are labeled A, B, C. For the larger panel, the three points were targets number 101, 110 and 1701; and for the smaller panel these were number 101, 108 and 1301. For the larger panel, the Y-axis is parallel to the line joining points 101 and 110, and the X-axis runs parallel to the line joining points 110 and 1710. Hence, most of the points on each panel show negative Z coordinate values.

## VI. Leveling Screw Locations

Three targets (A, B, C) out of four in Fig. 10 were chosen-adjacent to the panel-antenna backup truss adjustment leveling screws. In measuring points on the panel and in adjusting/leveling the antenna panels, the crew used a precision level with 10 seconds of arc per 2 mm dial bubble, prior to PGM. (More modern optical instruments incorporate an optical pendulum that automatically levels the sighting axis of the instrument). In an air-conditioned space, the Wild N3 instrument was sighted upon a vertical invar leveling rod placed in physical contact with the targets. The rounded end of the rod rested upon the top surface of the black overlay of the target.

Tables 4 and 5 show the Z departures from the reference plane (taken through three corner targets of the panels). Note from Table 5 that the fourth target D (No. 1710) lies 0.0078 inch (0.198 mm) above the reference plane formed by targets A, B, and C. Similarly, from Table 4 on the small panel No. 3, the fourth target D (No. 1308) adjacent to the fourth leveling screw was -0.0024 inch (0.061 mm) from the reference plane (formed by targets A, B and C).

Suppose that the reference plane for a panel had been relocated, to minimize the coplanar errors of four corner points (instead of three points on each panel), then, the Z coordinate offsets of the larger panel No. 9 would appear to be within  $\pm 0.0020$  inch (0.050 mm) at each corner; the Z-numbers changed but the panel did not. The above values suggest that the measuring accuracy resulting from the use of the Wild N3 level and the invar rod approaches  $\pm 0.002$  inch (0.050 mm) at short distances.

## VII. Test Results

The Simulator computed the predicted accuracy prior to the field work as in Table 2. The actual PGM results for all targets and camera stations are given in Table 3. The accuracy ( $1\sigma$ ) of all coordinates of both panels was better than 0.013 mm (0.0005 in.). This corresponds to a precision ratio relative to the object diameter between 1/230,000 to 1/270,000. Beginning at a minimum number of four camera stations, the accuracy of PGM self-calibrated triangulation improves directly as the square root of the number of rays (or lines) used to determine the location of each point/target, hence Table 6 is formed. Accuracies of 1/350,000 are possible, theoretically, with 16 rays per target.

Note that to demonstrate an accuracy better than 1 part in 350,000 would have required consideration of the thickness of each target (and in turn the additional thickness of the black overlay), the obliquity of the ray from the normal to the panel surface, and the separation of the nodal planes within the camera lens. Although the terms of these additional equations are straight-forward, there would have been additional computational expenses. Note also that theodolite accuracy peaks at a precision ratio of 1 part in 64,000 at best.

The two antenna panels are part of a contoured quasi-paraboloid (shaped) surface which usually requires 12 to 15 terms in a power series that characterize the main reflector surface. Appendix A is provided for reference by users who may wish to proceed with best-fitting of the data in order to determine RF path lengths, surface RMS, etc., for both parabolic or quasi-parabolic surfaces.

In addition, the riveted panel specifications were required to be checked with tooling accurate to  $\pm 0.076$  mm (0.003 in.). By using this pilot PGM demonstration, the two panels were measured to accuracies that were six times better.

## VIII. Future Work

Can the "not precisely levelled" panels coordinate data be transformed into the data that would have been recorded if the adjustment corner points (four to six points) had been perfectly positioned prior to the PGM test? If so, can an antenna panel be removed from the shipping crate when received at a site, "thrown" on the ground, photographed and the leveling done in a computer (instead of a sweep template jig) to detect possible damage during transport?

These thoughts could be used as the impetus for future work in software development to detect panel corner warping or panel stressing problems. Such capability should lead to a

better understanding of panel deflections and performance under loads, (Ref. 6) or the little traumas of shipping.

If adequate accuracy can be obtained in this process of converting photogrammetric panel data, the future might include such possibilities as:

(1) Obviate the use/need for quality assurance (QA) inspection jigs and sweep templates.

- (2) Eliminate the expensive labor paid for the adjustment of the panels in the jigs.
- (3) Shorten scheduling spans.
- (4) Provide a measuring accuracy of better than 0.025 mm (0.001 inch) that is only 20 percent of the 0.13 mm (0.005 inch) high quality panel tolerance (rms) requirements planned for future antennas.

## References

1. Brown, D. C., *The Simultaneous Determination of the Orientation and Lens Distortion of a Photogrammetric Camera*, RCAMTP Data Reduction Technical Report No. 33, Patrick Air Force Base, Florida, 1956.
2. Brown, D. C., *Evolution, Application and Potential of the Bundle Method of Photogrammetric Triangulation*, International Society for Photogrammetry, Commission III, Symposium in Stuttgart, W. Germany, Sept. 1974.
3. Brown, D. C., *Application of Close Range Photogrammetry to Measurements of Structures in Orbit*, Geodetic Services Inc. (GSI) Technical Report No. 80-012, Vols. I and II, prepared under contract to Rockwell International, Space Systems Group, Dept. 379, Downey, CA, September 1980.
4. Brown, D. C., *STARS (Simultaneous Triangulation and Resection System) A Turnkey System for Close Range Photogrammetry*, Geodetic Services Inc. (GSI) Technical Report No. 82-007, March 10, 1982.
5. Fraser, C. S., *Photogrammetric Measurement of Microwave Antennas*, Geodetic Services Inc. (GSI) Technical Report No. 86-001, May 1986.
6. Chian, C. T., and Levy, R. "Load-Deflection Tests and Computer Analysis of a High-Precision Adhesive-Bonded Antenna Reflector Panel," *TDA Progress Report 42-82*, pp. 68-81, Jet Propulsion Laboratory, Pasadena, Calif., August 1985.

**Table 4. Sample results for small panel No. 3**

Day No.	Description
1	Panels aligned on floor jigs, and targets placed.
2	2-man crew arrived on site. Photography of two panels. Developed and checked the films for four networks. Packed equipment and departed site (10-hr stay)
3	Film shipped by air.
4	Film arrived at GSI in Melbourne, Florida. Coordinates were read on the automated film comparator.
5	Verified the reference planes for the data. The computations were completed, the data plotted the same day.

Point No.	Coordinates, inch		
	X	Y	Z
1	-37.1161	38.6390	0.1638
2	-31.3560	-8.8819	-0.5642
3	-27.1867	-42.4212	0.3118
101	-35.7950	-35.1000	0.0000 <sup>a</sup>
102	-34.6751	26.1338	-0.2728
108	-27.0827	-35.0917	0.0000 <sup>a</sup>
201	-29.2056	32.8568	-0.3791
1208	20.2674	-36.7120	-0.1152
1301	35.7950	35.1000	0.0000 <sup>a</sup>
1308	26.7389	-35.1347	-0.0024 <sup>b</sup>
1601	-40.8885	41.7420	-0.0902

<sup>a</sup>Points 101, 108 and 1301 make the reference plane.<sup>b</sup>Point 1308 (not used in datum) shows a 0.0024 inch offset.**Table 2. Predicted PGM accuracy as given by Simulator output for large panel No. 9**

Number of Stations	Sigmas (rms), mm (inch)	Coordinate
4	0.028 (0.0011)	X
4	0.028 (0.0011)	Y
4	0.033 (0.0013)	Z

**Table 5. Sample results for large panel No. 9**

Point No.	Coordinates, inch		
	X	Y	Z
1	-42.6167	53.1027	0.11152
101	-41.5000	48.7352	0.0000 <sup>a</sup>
102	-41.4862	38.5145	-0.1795
110	-41.5000	-48.7352	0.0000 <sup>a</sup>
1701	-41.3086	-44.4043	0.0000 <sup>a</sup>
1710	31.2793	-52.5865	0.0078 <sup>b</sup>
2007	-9.9926	-49.8801	-0.3225

<sup>a</sup>Points 101, 110 and 1701 form the reference plane.<sup>b</sup>Point 1710 (not used in datum) shows a 0.0078 inch offset.**Table 3. Actual PGM test accuracy**

Panel No.	Number of Stations	Sigma (rms), mm (inch)	Precision Ratio <sup>a</sup> , sigma/object diameter
3 (small)	8	0.013 (0.0005)	1/230,000
9 (large)	8	0.013 (0.0005)	1/270,000

<sup>a</sup>This is taken as 2.91 m (9.55 ft) for the small panel diagonal and 3.47 m (11.40 ft) for the large panel diagonal.

**Table 6. Theoretical variation of PGM accuracy as a function of number of stations**

No. of Stations, or Lines per Object	1-Sigma (rms), mm (inch)	Precision Ratio <sup>a</sup> , sigma/object dia.
8	0.013 (0.00050)	1/250,000 (actual)
4	0.018 (0.00071)	1/177,000
16	0.009 (0.00035)	1/350,000

<sup>a</sup>Based on an average panel diagonal of 3.2 m (10.4 ft).

ORIGINAL PAGE IS  
OF POOR QUALITY

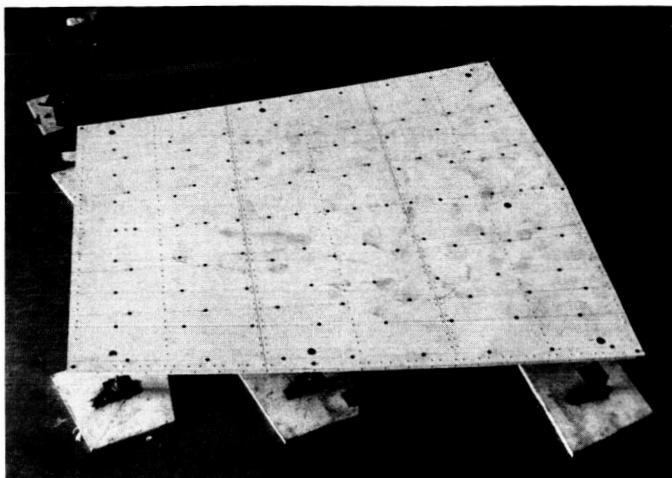


Fig. 1. Small panel No. 3 with PGM targets

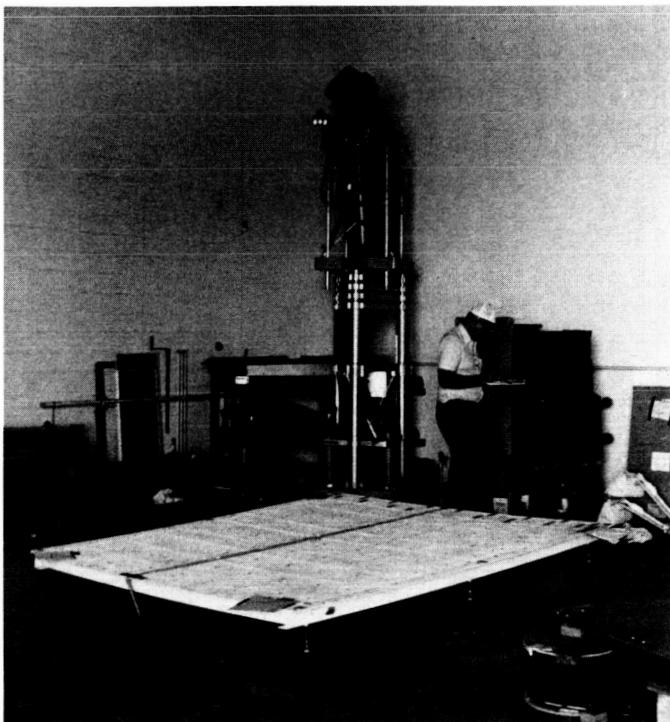


Fig. 3. Camera position relative to panel No. 9

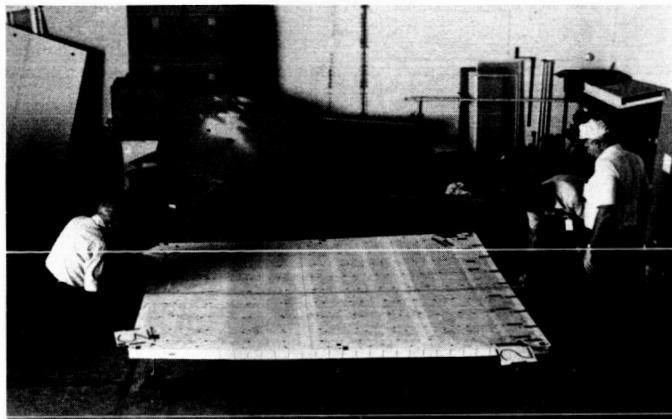


Fig. 2. Installation of scale tape on large panel No. 9

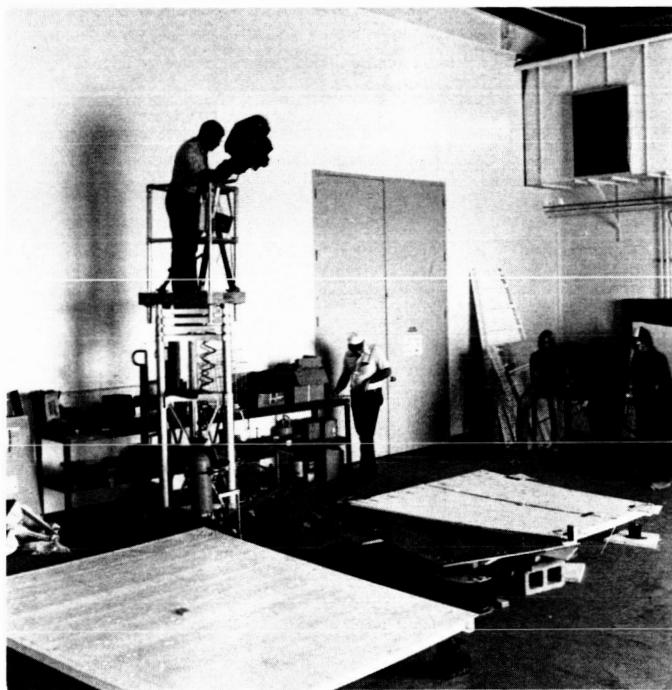


Fig. 4. Two panels ready for photography

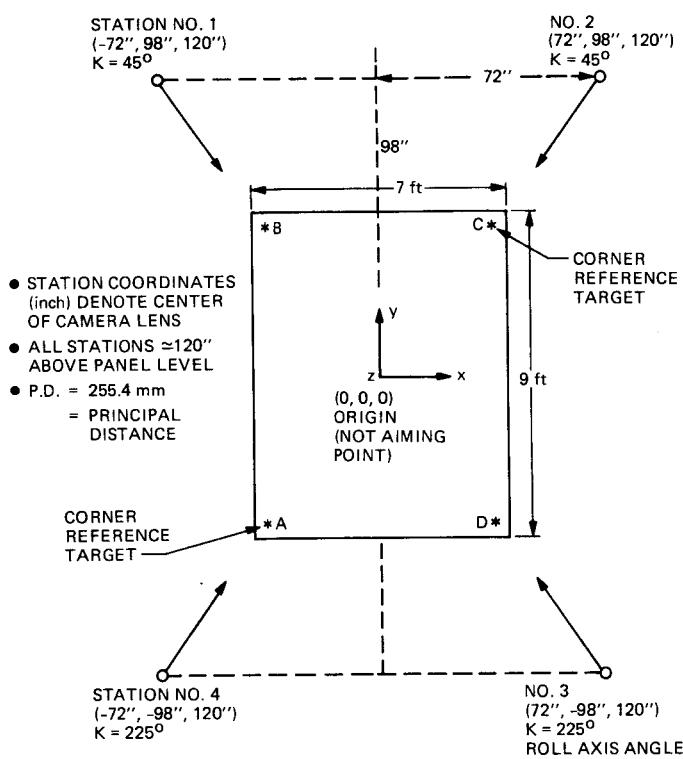


Fig. 5. Camera station locations relative to large panel No. 9

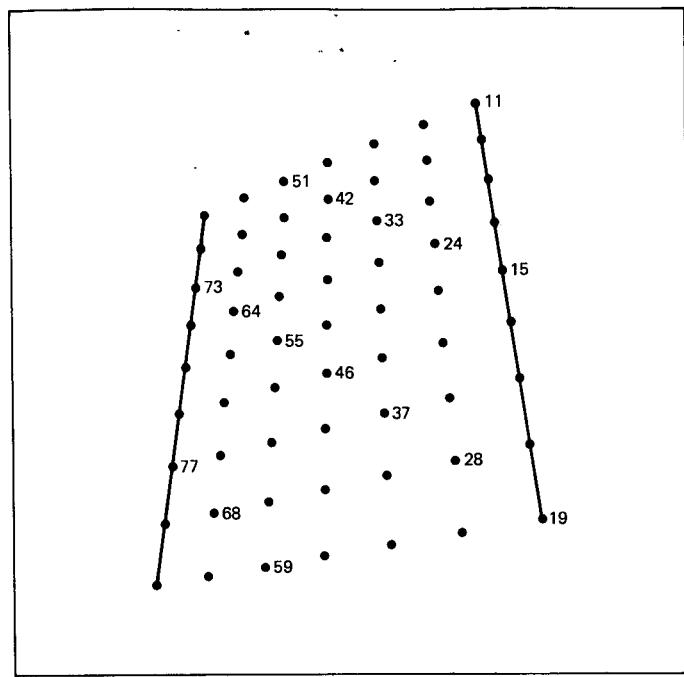


Fig. 6. Simulator output for camera station No. 1

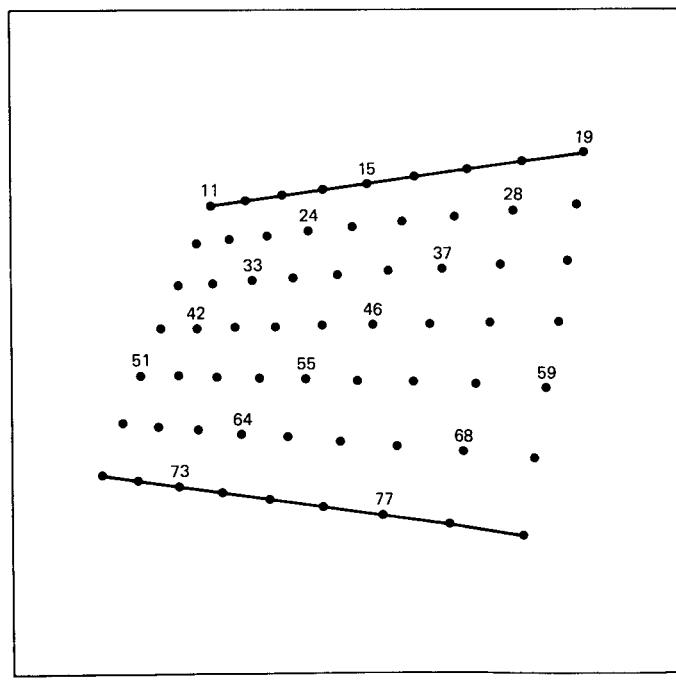
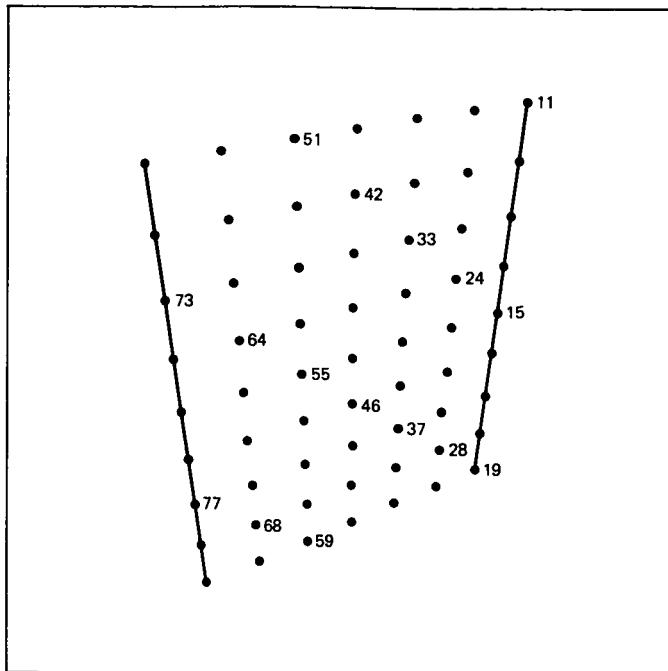
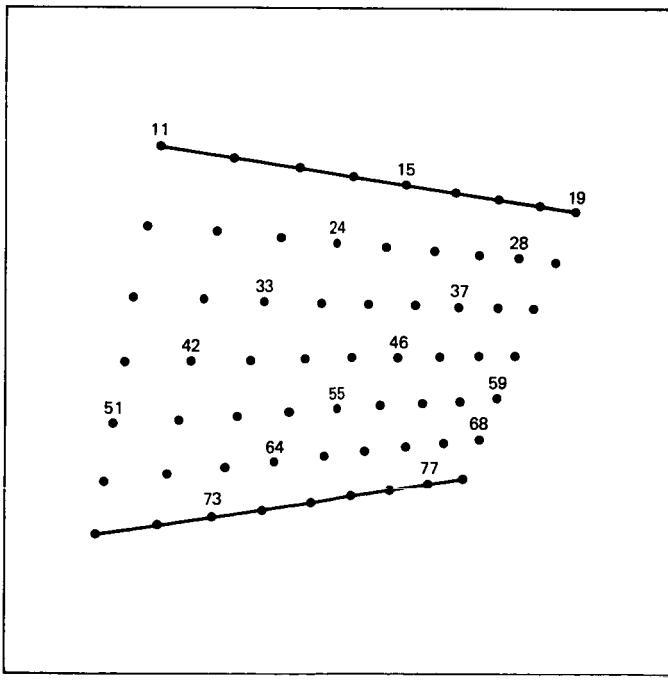


Fig. 7. Simulator output for camera station No. 2



FINAL FILM DIMENSION 23 X 23 cm

**Fig. 8. Simulator output for camera station No. 3**



FINAL FILM DEMENSION 23 X 23 cm

**Fig. 9. Simulator output for camera station No. 4**

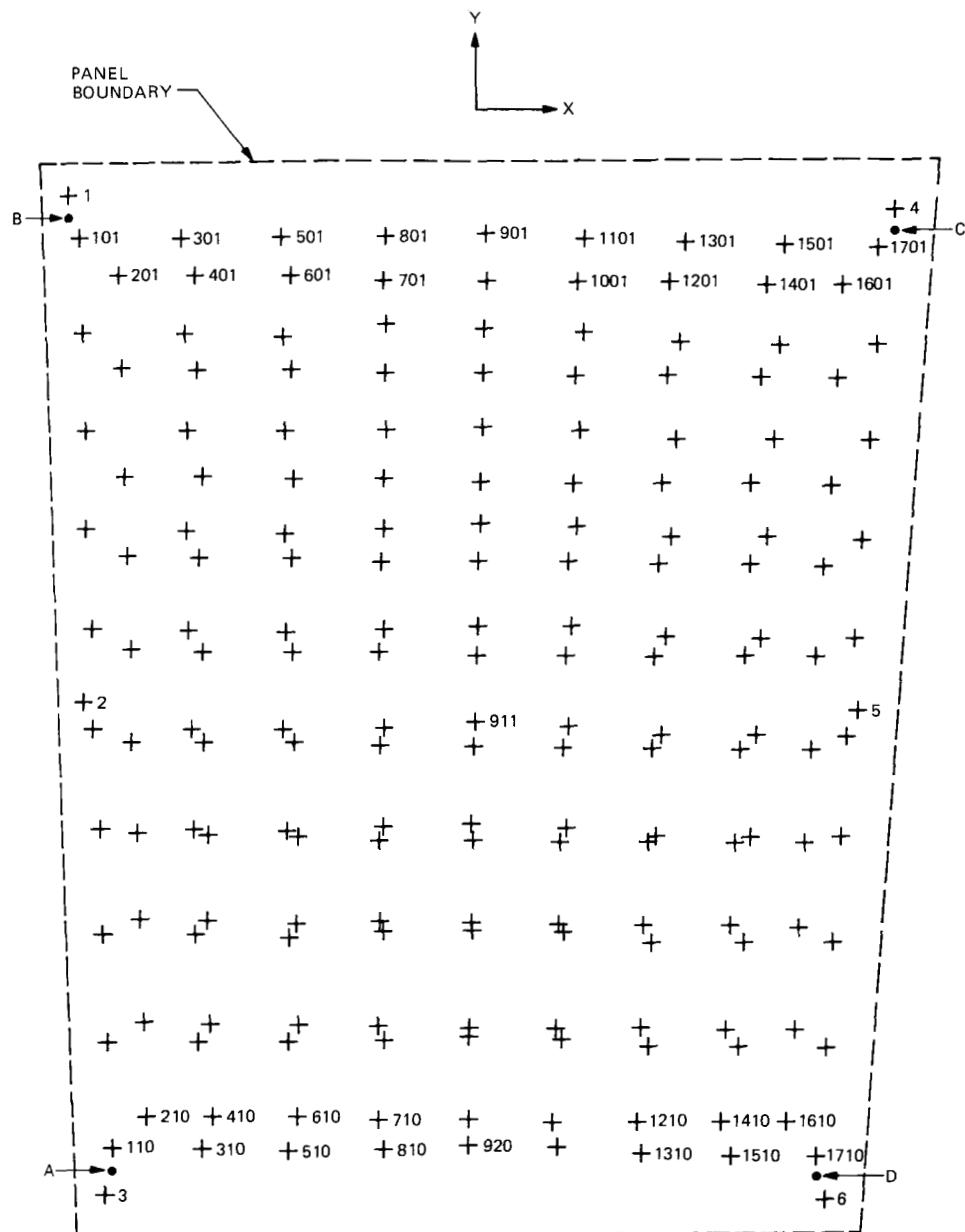


Fig. 10. Target distribution for large panel No. 9

## Appendix A

### Model for Best-Fit Parabola Computation to PGM Data

#### I. Model for Ideal Paraboloid

The basic equation of a paraboloid of revolution, as shown in Figure A-1, is:

$$x_p^2 + y_p^2 = 4fz_p$$

or

$$x_p^2 + y_p^2 + z_p^2 = 4fz_p + z_p^2 \quad (\text{A.1})$$

Let

$$\xi = \begin{bmatrix} x_p \\ y_p \\ z_p \end{bmatrix} \quad (\text{A.2})$$

hence,

$$[\xi]^T [\xi] = 4fz_p + z_p^2 \quad (\text{A.3})$$

The objective is to develop transformation equations between the measured  $(x, y, z)$  photogrammetric coordinates (with arbitrary origin and orientation) and best-fit paraboloid of revolution coordinates  $(x_p, y_p, z_p)$  with origin at the vertex. Transformation is achieved through a vertex translation to  $(x_0, y_0, z_0)$  and by two axial rotations ( $\alpha$  and  $\omega$ ).

$$\begin{bmatrix} x_p \\ y_p \\ z_p \end{bmatrix} = \begin{bmatrix} -\cos \alpha & \sin \alpha & 0 \\ -\sin \alpha \sin \omega & -\cos \alpha \sin \omega & \cos \omega \\ \sin \alpha \cos \omega & \cos \alpha \cos \omega & \sin \omega \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{bmatrix} \quad (\text{A.4})$$

or

$$[\xi] = [R] [\eta]$$

Substitution of Equation (A.4) into (A.3) yields:

$$[\eta]^T [R]^T [R] [\eta] = 4fz_p + z_p^2$$

because  $R$  is unitary orthogonal

$$[R]^T [R] = [I]$$

hence,

$$\eta^T \eta = 4fR_3 \eta + (R_3 \eta)^2 \quad (\text{A.5})$$

where  $R_3$  is the third row of  $R$  given by:

$$R_3 = [D \ E \ F] = [\sin \alpha \cos \omega \ \cos \alpha \cos \omega \ \sin \alpha] \quad (\text{A.6})$$

Equation (A.5) can now be expanded into the form:

$$\begin{aligned} 0 &= (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 \\ &- 4f[D(x - x_0) + E(y - y_0) + F(z - z_0)] \\ &- [D(x - x_0) + E(y - y_0) + F(z - z_0)]^2 \end{aligned} \quad (\text{A.7})$$

In computing a best-fit surface the six parameters  $x_0$ ,  $y_0$ ,  $z_0$ ,  $\alpha$ ,  $\omega$ , and  $f$  need to be computed. These parameters, along with the measured photogrammetric  $(x, y, z)$  coordinates make up the terms of Equation (A.7). Every point on the surface generates a separate equation.

Finding the best-fitting surface characteristics entails a solution of the parameters of Equation (A.7) via the method of least-squares through an iterative process with successive estimates of the 6 parameters. In linear form, a full observation equation set corresponding to Equation (A.7) can be written as follows:

$$[B] [r] + [A] [\delta] + [w] = 0 \quad (\text{A.8})$$

where

$[B]$  = sparse matrix of partial derivatives with respect to  $(x, y, z)$  coordinates computed at each point  $n$ . Dimensions  $(n \times 3n)$

$[r]$  = vector of coordinate residuals;  $r^T [r_x, r_y, r_z]$ ,  $\dots, r_{xn}, r_{yn}, r_{zn}$ . Dimensions  $(3n \times 1)$

$[A]$  = matrix of partial derivatives with respect to the parameters  $x_0, y_0, z_0, \alpha, \omega$  and  $f$ . Dimensions  $(n \times 6)$

and

$[\delta]$  = least-squares corrections to parameter estimates to yield most probable solution. Dimensions  $(6 \times 1)$

$$\delta^T = [\delta x_0, \delta y_0, \delta z_0, \delta \alpha, \delta \omega, \delta f]^T$$

and

$[w]$  = value of Eq. (A.7) using measured  $(x, y, z)$  coordinates and using estimates for parameter  $x_0, y_0, z_0, \alpha, \omega$ , and  $f$ .

where  $n$  is the number of photogrammetric points.

The solution to the overdetermined system, Equation (A.8), is given by:

$$[\delta] = -[A^T(BP^{-1}B^T)^{-1}A]^{-1}[A^T(BP^{-1}B^T)^{-1}w] \quad (A.9)$$

where  $P$  is the weight matrix corresponding to the variances of the  $x, y, z$  coordinates.

The accuracy estimates ( $\sigma$ , standard error value) for the parameters are obtained as:

$$C_\delta = \begin{bmatrix} \sigma_{x_0}^2 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \sigma_{y_0}^2 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \sigma_{z_0}^2 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \sigma_\alpha^2 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \sigma_\omega^2 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \sigma_f^2 \end{bmatrix}$$

$$= \sigma_0^2 (A^T(BP^{-1}B^T)^{-1}A)^{-1} \quad (A.10)$$

where  $\sigma_0^2$  is the variance factor.

The solution for  $x_0, y_0, z_0, \alpha, \omega$  and  $f$  which is an iterative procedure minimizes the quadratic form  $r^T Pr$  where:

$$r^T Pr = \sum_{i=1}^n \left( \frac{r_x^2}{\sigma_{x_i}^2} + \frac{r_y^2}{\sigma_{y_i}^2} + \frac{r_z^2}{\sigma_{z_i}^2} \right) \rightarrow \text{minimum} \quad (A.11)$$

Here,  $\sigma_{x_i}^2, \sigma_{y_i}^2$ , and  $\sigma_{z_i}^2$  are coordinate standard errors.

Once the six parameters of the best-fit-parabola are computed, the corresponding point on parabola  $(x_p, y_p, z_p)$ , coordinates are computed from Equation (A.4). The  $z$ -departure is then defined as:

$$Dz = z_p^{(\text{measured})} - [(x_p^2 + y_p^2)/4f] \quad (A.12)$$

The RMS value for the surface is computed by:

$$\text{RMS of } Dz \text{ values} = [\sum (Dz^2)/n]^{1/2} \quad (A.13)$$

where "n" is the number of photogrammetric points.

## II. Model for Quasi Paraboloid (e.g., Shaped Surfaces)

Initially,  $x_0, y_0, z_0, \alpha, \omega$ , and  $f$  are computed according to the above model. Because the surface is not a parabola, however, the  $z_p$  origin of the parabola coordinates is not strictly defined ( $f$  is also not applicable). Given a shape profile for the surface in the form of:

$$z'_p = f(r) \quad (A.14)$$

the  $z_p$  origin is defined such that

$$\sum_{i=1}^n (z_p - z'_{p_i}) = 0 \quad (A.15)$$

That is, the sum of all the final computed  $z$ -departure is zero. The computed offset to  $z_p$ , namely  $\delta z_p$ , is then applied to the parabola  $z_p$  coordinates such that Equation (A.15) is satisfied. The surface RMS value is then given as

$$\text{RMS of } Dz_p \text{ value} = \left\{ \sum_{i=1}^n \frac{[z_{p_i}^{(\text{final})} - z'_{p_i}]^2}{n} \right\}^{1/2} \quad (A.16)$$

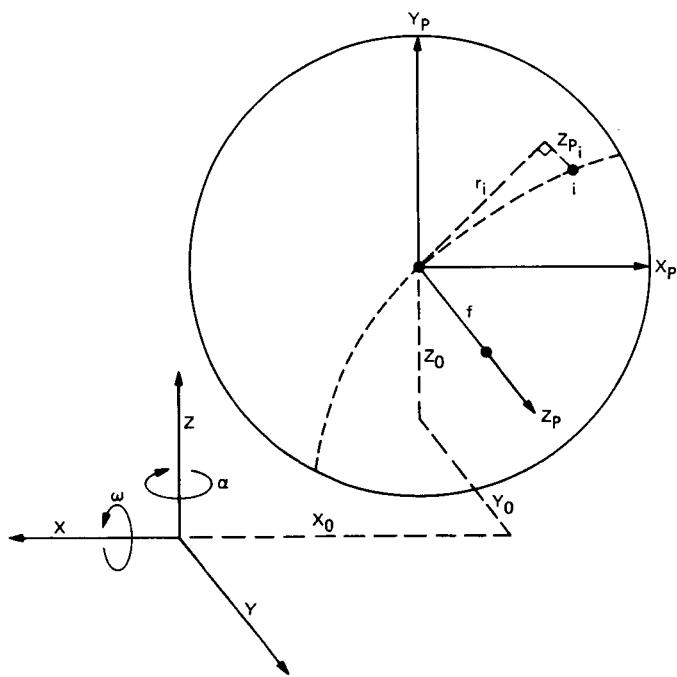


Fig. A-1. Coordinates for best-fit paraboloid

D23-32

22

QH2

# Pointing Calibration of the MKIVA DSN Antennas Voyager 2 Uranus Encounter Operations Support

R. Stevens

Office of Telecommunications and Data Acquisition

R. L. Riggs

TDA Engineering Office

B. Wood

Bendix Field Engineering Corporation

*The MKIVA DSN introduced significant changes to the pointing systems of the 34-meter and 64-meter diameter antennas. To support the Voyager 2 Uranus Encounter, the systems had to be accurately calibrated. Reliable techniques for use of the calibrations during intense mission support activity had to be provided.*

*This article describes the techniques used to make the antenna pointing calibrations and to demonstrate their operational use. The results of the calibrations are summarized.*

## I. Introduction

In October 1985, a concentrated effort was undertaken to complete and demonstrate the antenna pointing calibrations needed to support the January 1986 Voyager 2 Uranus Encounter (UE) operations. There were two principal objectives of the effort.

The first objective was to develop calibrations of the systematic pointing errors for all MKIVA DSN antennas as required to support Voyager 2 UE operations. Interim operational techniques for making and using the calibrations were to be developed and promulgated.

The second objective was to demonstrate the technical and operational pointing capabilities of the antennas in a realistic Voyager 2 mission support environment.

Both objectives were accomplished.

This article describes the MKIVA antenna pointing system, and the techniques used to calibrate its major systematic errors. The results obtained from the calibrations are summarized.

## II. Need for Antenna Calibrations

The MKIVA antenna pointing system was a new system requiring new calibrations and new procedures for obtaining the calibrations. The inheritance from the MKIII era was generally inapplicable. The equipment and detailed methods for producing the antenna pointing predicts were changed. At the Network Support Subsystem in the JPL Network Operations Center, the old multiuser computers used to sup-

port predict generation were replaced by new VAX dedicated computers. The programs in the new computers use new algorithms and new operating procedures.

At the network complexes, all pointing system equipment, with the exception of the 64-m and 34-m Ha-Dec antenna structures and drives, was replaced. New 34-m Az-El high-efficiency antennas were implemented at the Goldstone and Canberra complexes. The old Scientific Data Systems antenna pointing computer was replaced with a ModComp Classic central computer plus a microcomputer located at each antenna. The antenna control systems were completely redesigned. In the MKIVA configuration, microcomputers at each antenna enable monitor and control of the antenna from a remote operating console in the Signal Processing Center.

### **III. Technical Requirements for Pointing Calibrations: Operational Requirements for Use of Calibrations**

Voyager 2 UE support required the capability to point the DSN antenna beams solely by predict control very close to the direction of arrival of the signal from the spacecraft (S/C). "Very close" means that the error in pointing would not significantly degrade the received signal. Predict control implies "blind pointing" — pointing without the benefit of knowledge of the received signal strength or any other indicator for operator guidance.

Accurate blind pointing was specifically required to support the Voyager 2 UE critical radio science data acquisition, because without it the data would be degraded. Also, blind pointing was required for S/C acquisition and reacquisition, and for emergency command.

The pointing accuracy specification adopted for the Voyager 2 UE preparations was 0.16 of the antenna half-power beamwidth. That corresponds to  $\leq 6$  and  $\leq 11$  millidegrees (mdeg) total beam pointing error for the 64-m and 34-m antennas respectively at X-band (8.4 GHz). That specification provides  $\leq 0.3$  dB signal degradation from pointing errors.

The requirements we adopted for operational use of the antenna calibrations were that the process be practical and reliable in the station environment during Voyager 2 UE support activities. During the test and demonstration work we tried to develop and validate a system capability and the operating procedures to meet those requirements. It soon became evident that automatic antenna systematic error

correction capability was essential. Manual pointing offset entry at the Link Monitor and Control console during a high activity period was not viable.

The needed automatic capability was planned for the MKIVA, and strongly endorsed by the operational complexes. However, it was not available at the beginning of our work. A determined effort by engineering and operations personnel produced the automatic capability in time to support our real needs.

### **IV. The MKIVA DSN Antenna Pointing System**

The basic block diagram of the end-to-end antenna pointing system is shown in Fig. 1; the glossary of abbreviations used is in Table 1. The process of pointing a MKIVA DSN antenna at a S/C signal is as follows.

At JPL, the Project provides S/C ephemeris data, in the form of position and velocity vectors, to the NSS Navigation team. The data are on magnetic tape. Using VAX computers, these data are converted into topocentric antenna pointing predicts for each station and placed on file in the NSS. The predicts give direction cosines of the S/C at two-minute time intervals. Several days before a scheduled tracking period, the predicts are transmitted over the NASCOM high-speed data lines via the complex SPC ARA to the CMC, where they are filed on disk. Before each tracking pass, the complex CMC operator transfers the antenna predicts to the APA. The APA transforms the direction cosines of the predicts to Az-El coordinates. The LMC operator then downloads the antenna Az-El predict points and a pointing systematic error correction table from the APA computer disk to the antenna mounted ACS microcomputer.

The ACS automatically corrects the pointing predicts for refraction and subreflector position, and adds the proper systematic error correction and any manually entered antenna offsets. Then, the ACS interpolates the predict points into an angular position command for each antenna axis once per second. These ACS position commands are used by the ASC or the 64-m MEC to generate rate commands that are subsequently used to point the antenna. In this mode, the antenna is being pointed automatically in response to the predicts with automatic corrections. It is being "blind pointed."

The antenna can also be pointed in a mode that can sense and correct for all pointing errors including those from the S/C predicts supplied. That mode is Conscan. It can correct for errors as large as the half-power beam width of the antenna. Conscan (for conical scan) works as follows. Again, refer to Fig. 1. The APA generates a circular scan pattern for the

antenna and sends it to the ACS. The ACS adds the scan pattern to the corrected pointing angle predicts. Software in the REC computes and then sends received signal levels to the APA via the SPC LAN. The correlation of the scan position of the antenna, as reported by the ACS, with the received signal level variations, allows the APA to compute offset changes to the scan pattern center. The APA sends the offset changes to the ACS. Thus, within the capability of the closed-loop control system, the scan center is pointed precisely to the apparent direction of arrival of the S/C signal.

## V. Pointing Error Calibration Concept

The process of calibrating antenna systematic pointing errors involves iteration of three steps. Step 1 is to collect pointing offset data from observations of radio sources of accurately known position. Step 2 is to use the pointing offset data to determine the constant parameters in a model that represents the systematic pointing error behavior of the antenna. Step 3 is to use the model to generate a table of antenna systematic pointing errors vs antenna pointing direction.

The three-step process is iterated until a final table of acceptable accuracy is obtained. The final systematic error table is used to correct the pointing of the predict driven antenna in operational tracking support.

The complete process as used for the Voyager 2 UE antenna calibrations is depicted in Fig. 2. It is described in the following paragraphs. Again, refer to the Glossary of Abbreviations in Table 1. Also, refer to the Appendix for a summary of the software programs used.

### A. Step 1: Collection of Pointing Offset Data

**1. Offset Data from Radio Star Observations.** A process developed by the Madrid Deep Space Communication Complex (MDSCC) was used for initial pointing calibration of the MKIV-A configured antenna using radio stars. It was used because initially we lacked confidence in the accuracy of the new S/C antenna pointing predict system.

The basic technique is to scan a strong and precisely located radio star and observe the received signal power level as a function of scan position about the star's predicted location. The functional assemblage of the antenna pointing data acquisition and analysis programs used is shown in Fig. 3. The observation on a preplanned set of stars, and the data collection and analysis to determine pointing offsets vs azimuth and elevation can be accomplished essentially automatically. That capability is provided using programs written at MDSCC. Calibrated signal power from a radio star is obtained using an

updated DSN MKIII NAR program in the MDA computer. NAR diodes at the antenna front end are modulated by the MDA computer, and the receiver IF output is square-law detected. The detected signal is converted to a varying frequency, processed by the MDA frequency counters, and recorded on disk.

After a series of radio star observations, the data are reduced to pointing angle offsets in azimuth and elevation using the MDSCC AGA program. The resulting offset data are input to the PHO program (see Fig. 2) which produces an updated antenna systematic pointing error model. The systematic pointing error model development and the use of the PHO program are described in Subsection V.B.2.

Antenna pointing error models developed from the radio star observations were fairly accurate in sky areas near the stars observed. Typical accuracies were  $\leq 4$  mdeg for the 64-m antenna,  $\leq 6$  mdeg for the 34-m Az-El antenna and  $\leq 10$  mdeg for the 34-m Ha-Dec antenna.

Radio stars near the 23-deg south declination of the Voyager 2 S/C were not readily available. For that declination, Conscan offset data from Voyager 2 support passes were used to refine the error models derived from radio star tracks. The Conscan process is described in the next section.

**2. Offset Data from Tracking S/C in Conscan Mode.** The principle of obtaining pointing calibrations using Conscan is as follows. When the antenna is operating in Conscan mode, the apparent direction of the S/C radio signal is sensed. The difference between that direction and the direction the corrected pointing predicts define is the Conscan offset. The system resolves the offset into Az and El offset components that are recorded with the corresponding antenna Az and El, or Ha and Dec. To the extent that the Conscan system properly senses the apparent direction of the S/C signal, the Conscan offsets determine the total system pointing errors.

To verify the Conscan performance, beam scans were performed during Voyager 2 tracks at the Goldstone Deep Space Communications Complex (GDSCC); they demonstrated that the Conscan-defined axis and the antenna beam axis were coincident within the experimental error of 1 to 2 mdeg.

Any errors in the pointing predicts are embedded within the total system errors determined by the Conscan data. In the early phase of the calibration work there was a 3 to 5 mdeg error identified in the NSS predict calculations. That was repaired. Later, another predict error of 1 to 4 mdeg, depending on the complex, was found; it resulted from incorrect use of station location coordinates in the calculations. That error was not repaired during our work; however it was

stable, and was accounted for in the pointing calibrations. By the end of the work, we concluded that the NSS-supplied Voyager 2 pointing predicts were consistent to our level of visibility,  $\leq 1$  mdeg.

The computer-based process that was developed and used for collection and analysis of Conscan tracking offset data is depicted in Fig. 4. It is explained in the following paragraphs. We note that at the start of the calibration effort, the software programs to realize Fig. 4 did not exist. For example, Conscan tracking data were initially hand tabulated from the LMC console displays — a very impractical operational process. But in a most timely way, R. Livermore of the Canberra Deep Space Communications Complex (CDSCC) developed an IBM-PC compatible program to automatically log the needed antenna pointing data on a noninterference basis during mission support passes. That program, CAPTURE.BAS, provided an essential capability; it was modified, adapted, and used by all three complexes during the calibration effort.

To use the automatic logging program, an IBM-PC computer is connected to a monitor port on the APA modem patch panel (see Fig. 4). This accesses antenna information being sent to the APA from the ACS. Antenna pointing angles, azimuth and elevation Conscan offsets, and time are logged on diskettes and printed out on a line printer. Figure 5 is a sample of the collected data from CAPTURE.BAS.

The Conscan offset data collected are analyzed using a plotting program, PLOT.BAS, initially developed by R. Murray of the CDSCC (see Fig. 4). The program graphs the azimuth and elevation conscan offsets for the antenna as a function of azimuth (or local hour angle for the 34-m Az-Dec antennas). Another program, CSN-ANAL.EXE, by R. Riggs of JPL, provides the mean and standard deviation of the offsets, and other summary parameters of the antenna pointing performance.

The plotting and analysis programs were modified, adapted, and used by all complexes, and at JPL. Several samples of the outputs from the programs will be presented near the end of this article, in Section VI.

The Conscan offset data were used via the PHO program to upgrade the parameters in the antenna pointing systematic error model. The development of the model and the use of the PHO program are described in Subsection V.B.2.

Conscan offset data were collected regularly on an essentially noninterference basis from Voyager 2 pre-encounter period support passes. The data provided the basis for establishing accurate systematic pointing error correction to prepare all network antennas for supporting the Voyager 2 UE operations.

## B. Step 2: Use of the Antenna Systematic Pointing Error Model

**1. Introduction – Description of the Model.** The concept of the antenna systematic pointing error model and the rationale of its use are discussed in the next paragraphs. Excepting environmental effects, the major sources of errors in an antenna pointing system are systematic and repetitive, and can therefore be closely modeled. Examples are residual errors in the geometric alignment of the mount axes and shifting of the antenna beam relative to the elevation axis angle readout as the antenna is tipped.

An antenna error model is used in the pointing calibration process. In the process, data from S/C or radio star observations are used to establish the parameters in the model; the model is then used to generate a systematic error correction table for accurately pointing the antenna.

The basic pointing error modeling approach used was originally devised by optical astronomers. Radio astronomers adapted the model for antennas. The model is based on logical expected physical behavior of the antenna. It has been successfully applied at major radio astronomy facilities: for example, the Bonn 100-m Az-El antenna (Stumpff, Ref. 1) and the Haystack 37-m Az-El antenna (Meeks et al., Ref. 2).

A. Rius, MDSCC, adopted the work of Stumpff and Meeks to develop a pointing error model for the DSS 63 64-m Az-El antenna operating in the computer control mode. W. Peters, CDSCC, developed the error model for the DSS 42 34-m polar mount (Ha-Dec) antenna. Engineering personnel from the three complexes and from JPL collaborated in merging the two DSN antenna models to provide a preliminary error model for the 64-m Az-El antenna pointing under Master Equatorial control.

The complete pointing error model for an antenna is a sum of individual error functions. The individual error functions typically have a constant multiplier and depend on variables that are accessible in the system, for example, Az and El. For example, Table 2 shows the individual error sources and the El and Cross-El error models for the DSN 34-m Az-El antennas.

The  $P$  constants in the individual error function models of Table 2 can be determined by collecting and analyzing data from instrumented pointing calibrations on radio signals whose sky positions are accurately known. The quality of the functional representations and their constants can be improved by collection and analysis of additional pointing calibration data.

We note that the systematic pointing correction for the Y axis motion of the subreflector is not included in the pointing error model of Table 2. That function is accomplished automatically by the ACS. The ACS receives subreflector position information from the SRC via the ASC. The required elevation-angle correction is computed and summed with other offsets in the generation of position data.

As previously stated, our use of a systematic error model for calibrating an antenna for blind pointing is iteration of a three-step process: (1) measurements of pointing errors on sources of known position, (2) determination of the error model parameters from the measurements, and (3) development of a table of required pointing offsets vs pointing angles from the completed error model. One might ask if such a complex process is necessary, or fundamentally useful.

In principle, the pointing errors derived from the S/C or radio star observations can be used directly to develop a systematic error table. However, the table so developed is accurate only in pointing directions very near the directions (i.e., within 2 to 4 deg) of the observations. Sans model, no legitimate extrapolation or interpolation of the errors is available; cum model, a compact set of properly chosen observations can yield an error table useful for the full sky coverage of the antenna. Also, the model process provides needed filtering of errors in the observational data.

That economy of calibration effort in an operational environment has fueled development, refinement, and use of the model technique at all major radio astronomy facilities. The DSN is developing error models of its operational antennas for the same reason.

Also, the model technique provides a useful, powerful tool to antenna design engineers. That is because the separable individual model error functions are identified with particular physical characteristics of an antenna. Careful analysis of pointing calibration data in relation to specific terms in the model can yield quantitative understanding of the stability and precision of elements of an antenna. That understanding is essential to the process of engineering improvements to the DSN antennas for their current applications, and for future applications (for example, ka-band operations on the 34-m and 64-m/70-m Az-El antennas).

We must observe that the models developed during the calibration effort were based almost entirely on ConsScan offsets relative to NSS pointing predicts from Voyager 2 tracking. The quality of the models in the sky regions of the Voyager 2 S/C (23 deg S. Dec) was demonstrated to be good. In other regions of the sky, the quality of the models was not examined; that remains a task for the future.

**2. Updating the Antenna Systematic Pointing Error Models from the Observational Data.** The observational data are used by the PHO error model program to generate or refine the antenna systematic pointing error model. For radio star observations, refer to the diagram of Fig. 3. The observational data are available from the MDA disk via the AGA program. The AGA output data must be translated to the proper format for input to the PHO error model program. That translation involves extensive hand editing and manipulation of the data.

For S/C ConsScan offset observations, refer to the diagram of Fig. 4. ConsScan offset data logged in the IBM-PC are hand edited to remove obvious bad data points and abnormalities. That file is then converted to a format directly usable by the PHO program. The conversion is done using CON2DAT, a utility program written by R. Livermore of CDSCC.

Refer back to Fig. 2, the diagram of the complete three-step calibration process. PHO reads the converted ConsScan or radio star observations offset data file. PHO then sums the offsets from the "original" error model and the offsets from the new observations (the "original" model is the model used to generate the offsets used during the new observations). Finally, PHO uses a linear least-square error fitting routine to generate an updated systematic error model for the summed offsets.

### C. Step 3: Generation of the Systematic Error Correction Table from the Antenna Error Model

The antenna systematic error model is used to generate a new or updated table of antenna pointing offsets. The process is as follows.

Again, refer to Fig. 2. The model is processed by the IBM-PC program APACRCTB to generate an APA protocol compatible systematic error table. Basically, the table is a matrix of azimuth and elevation errors vs azimuth and elevation at 5-deg increments. The table includes a cyclic redundancy code (CRC) check number to verify its integrity.

The IBM-PC is then loaded with the PCPLOT VT-100 terminal emulator program, which allows direct communication with the APA through its maintenance terminal port. Then, using the Batch Operating System software, the APA is configured to receive the systematic error correction table.

After successful transfer to the APA high-capacity disk, the table is available for downloading to the ACS for S/C tracking support. During the support, the ACS automatically interpolates the systematic error table entries, combines the resulting errors with the pointing predicts, and provides corrected pointing commands for use by the ASC or MEC.

## **VI. Samples of MKIVA DSN Antenna Pointing Performance Data**

The plots in Figs. 6(a) through 6(h) display the pointing accuracy of the antennas of the three complexes as prepared for Voyager 2 UE support using the techniques described. They depict Conscan El and Cross-El components of the pointing errors relative to the predicts, and the total beam error calculated as the square root of the sum of the squares of the components.

The plots were obtained directly from our data capture and data plotting and analysis programs during Voyager 2 operational support passes. The data were obtained at X-band (8.4 GHz). During the early phases of the work, we demonstrated that Conscan offset results authentically indicate blind pointing capability. For illustration, Fig. 7 shows a plot of received signal level during a Voyager 2 pass at DSS 14 in which blind pointing was alternated with Conscan tracking. The received signal level is not significantly affected by the pointing mode changes. At CDSCC, long periods of accurate blind pointing were required for Voyager 2 UE Radio Science support. Therefore, the CDSCC performed extensive demonstrations of this type prior to the encounter.

## **VII. Secondary Results From the Work**

In the course of the work, we conducted some tests, and collected and superficially analyzed some experimental data, that looked interesting – potentially important. But we lacked the time to explore any of them in depth. Those items are discussed in the following paragraphs.

### **A. Pointing Errors from the Sun's Atmosphere**

X-band Conscan offset data were obtained from DSS 14 immediately before and after the Voyager 2 solar conjunction in December 1985. Also, some data were obtained from DSS 15 following the conjunction. The data in Fig. 8 show the approximate total pointing jitter vs Sun-Earth-Probe angle for DSS 14. The available data from DSS 15 are also shown. These data help characterize antenna system performance when tracking near the sun.

### **B. Refraction Correction Quality**

The ACS is operated with a fixed “default” refraction correction for each antenna. The correction is based on mean annual barometric pressure, temperature, and humidity of the particular complex.

The ACS software is designed to accept in situ observed meteorological parameters from the station Meteorological Monitoring Assembly (MMA) via the APA. At the time of our

work, the MMA could not provide the needed data; also, the APA software could not handle it. As a result, the ACS was not able to react to locally observed temperature, atmospheric pressure, or humidity parameters that deviate from the default values. Figure 9(a) displays plots of pass-average elevation angle Conscan offsets versus day of year from Voyager 2 tracks on DSS 14 and DSS 15. The antennas are colocated at GDSCC. The plots show significant errors that are highly correlated. Similar plots of azimuth angle offsets, Fig. 9(b), show small errors without evident correlation. The results point strongly to refraction correction error as the cause of the elevation errors.

To probe the matter, crude local weather observations were obtained during acquisition of a small set of the antenna Conscan offset data. Predicted angle-of-arrival deviations based on the observed atmospheric refraction effects were compared with the measured Conscan elevation offsets. The comparisons showed reasonable agreement.

We conclude that for X-band (8.4 GHz) pointing of the 64-m antennas, the refraction correction needs to incorporate some level of local and timely assessment of atmospheric conditions. There are planned implementation upgrades of the APA software and the MMA. They will allow the APA to accept needed data from the local MMA. That will enable automatic updating of the ACS refraction correction. Our results support the importance of the MMA and APA upgrades.

The refraction correction upgrade should significantly improve the operational accuracy of the antenna pointing by predicts. Concurrently, elimination of refraction error will aid engineering studies that use tracking error data to identify intrinsic capability, and hence growth capability, of the 64-m antennas.

### **C. Wind Effects on Pointing Accuracy**

Several times during testing and demonstration of antenna pointing at GDSCC, unusual azimuth offsets were observed during strong winds. The errors were variable, usually positive in sign, and were observed at both DSS 14 and DSS 15. The tracks were from east to west, with a maximum elevation of 32 deg. The winds were generally from the southwest.

Conscan errors caused by wind are expected to appear as offsets into the prevailing wind. Thus, wind from the southwest would cause positive offsets.

The results noted suggest that useful data on the effects of wind on beam pointing accuracy can be obtained as a by-product of X-band Conscan S/C support tracking. In concept,

merely be prepared to collect wind conditions and Conscan offsets when very windy days occur.

#### D. Sidereal Mode Tests

The primary ACS operating mode for the Voyager 2 UE was the predict mode. A secondary ACS mode, sidereal, was tested with systematic error tables to demonstrate it as a backup capability in case of a failure in the predicts mode system.

The planetary mode, a preferred backup, could not be used. That was because the ACS software then available could not support the use of systematic error tables in the planetary mode.

To operate in the sidereal mode, the conventional planetary predicts routinely supplied to the complexes by the NSS are used. The predict points, spaced a day apart, are manually interpolated to provide right ascension and declination coordinates and rates. At the time of track start, the coordinates, rates, and time are entered into the ACS.

The sidereal mode demonstrations yielded the same antenna performance as the prime predicts mode provided. Figure 10(a) shows antenna Conscan offsets during a DSS 14 Voyager 2 track when the sidereal and predict modes were alternately used. Figure 10(b) shows the received signal level during that track. For the period 2030 to 2200 GMT, the antenna was blind-pointed in the sidereal mode.

Interim operating procedures were developed and promulgated so that all stations could invoke sidereal as a backup tracking mode during the Voyager encounter operations. With the system capability at the time, LMC Console operations in the sidereal mode were complicated — errors were easy to make.

D. Girdner, GDSCC, planned and conducted the tests of the sidereal mode. He also developed the interim procedures.

#### E. Conscan Snap-on Tests

Dynamic response tests were run for demonstration and to evaluate the transient behavior of the DSS 15 antenna pointing system in the Conscan mode as used during Voyager 2 tracks.

The tests involved turning Conscan off and then offsetting the antenna by 30 percent of its beamwidth. The received signal level was observed on a strip-chart recorder and the Conscan offsets were logged when Conscan was turned back on. The signal level and Conscan offset signatures allowed dynamic characteristics to be analyzed for response time,

damping factor, and amount of cross coupling between the azimuth and elevation axes.

Analysis of the signal-level recording from single-axis offset snap-on tests of DSS 15 is summarized in Table 3. The signatures of the Conscan offsets from the test are shown in Figs. 11(a) and (b). The Conscan single-axis offset behavior and the signal-level results of Table 3 are consistent.

The Conscan snap-on tests indicated that the operating parameters of the antenna were approximately correct and that the Conscan system was performing acceptably for Voyager 2 support.

D. Girdner, GDSCC, planned and conducted the snap-on tests.

#### F. Observations on Pointing Predicts

We noted in Subsection V.A.2 that errors in the S/C pointing predicts are embedded in observations of Conscan offsets. Two errors of significance to 64-m X-band (8.4-GHz) pointing were identified. The first was repaired; the second, which is related to use of station location coordinates, is currently being investigated. When the second error is corrected, its effect on the antenna pointing calibrations must be accommodated. The pointing predicts need to be an impeccable standard for antenna pointing in S/C mission support.

Also, we have observed that during S/C Conscan support passes, valuable pointing offset data can be obtained concurrently, on a noninterference basis. High-precision predicts are a prerequisite to the analysis of such data for the purpose of identifying and quantifying the intrinsic performance capabilities of the DSN antennas.

### VIII. Concluding Observations, Acknowledgements

The process of automatic collection and machine analyses of Conscan offset data provided a very effective means of calibrating and monitoring antenna system pointing. Usually the data were of excellent quality. Almost all of the Conscan data were collected during committed Voyager 2 mission support passes — had we not been able to do that, we could not have achieved our objectives in the time available.

We express our appreciation to the Voyager Project, and especially D. G. Griffith, the Flight Operations Office Manager, for cooperation and support of our tests, demonstrations, and collection of calibration data.

The JPL engineering and operations organizations provided exemplary support throughout the calibration activity. We cite

especially personnel of the Ground Antennas and Facilities Engineering Section, the Radio Frequency and Microwave Subsystems Section, the DSN Control Center Operations Section, and the DSN Operations and Engineering Support Section.

Generally, the calibration and demonstration techniques used were applied first at the GDSCC, a special burden on the staff during the busy period of preparations for the Voyager 2

UE. All of the complex personnel involved were patient, resourceful, and professional.

All of the DSN complexes made critical contributions to the techniques for performing the calibrations and the demonstrations. The cooperation and support from the Management and Staff of the CDSCC, MDSCC, and GDSCC were splendid. Also, and most importantly, they calibrated their antennas in a timely manner.

## References

1. Stumpff, P., *Astronomical Pointing Theory for Radio Telescopes*. Klein-heibacher Berichte, Vol. 15, Fornmolde-technischon Zentralamt, Darmstadt, pp. 431-437, 1972, (Translation available from the JPL Library under No. 78-000796 C.1).
2. Meeks, M. L., Ball, H. A., and Hull, A. B., "The Pointing Calibration of the Haystack Antenna," *IEEE Transactions on Antennas and Propagation*, Vol. AP-16, No. 6, pp. 746-751, November 1968.

**Table 1. Glossary of abbreviations for text and Figs. 1, 2, 3, and 4**

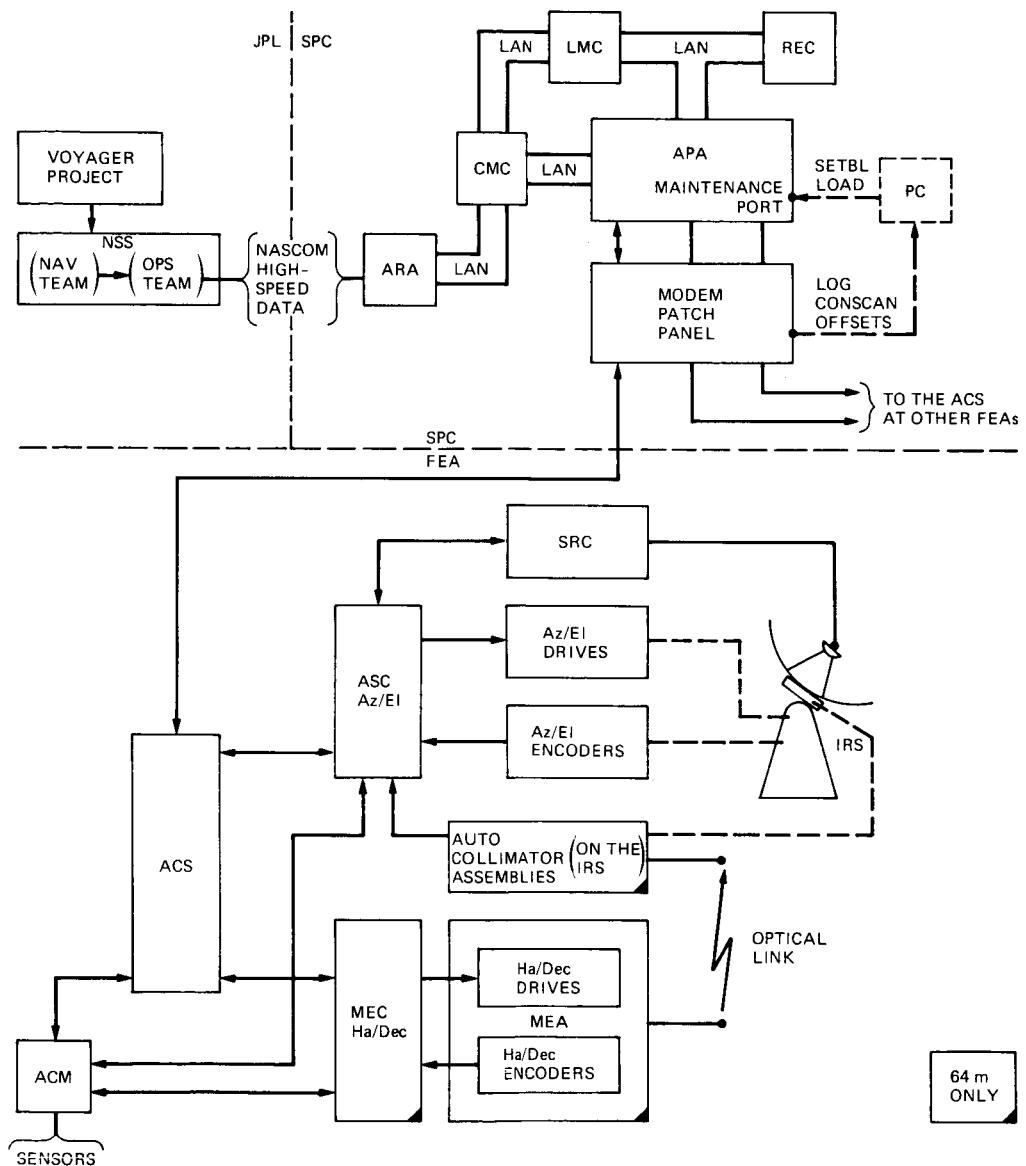
ACM:	Antenna Control and Monitor
ACS:	Antenna Control Subassembly
APA:	Antenna Pointing Assembly
ARA:	Area Routing Assembly
ASC:	Antenna Servo Controller
CMC:	Central Monitor and Control (console)
FEA:	front end area
IRS:	intermediate reference structure
LAN:	Local Area Network
LMC:	Link Monitor and Control (console)
MEA:	Master Equatorial Assembly
MEC:	Master Equatorial Controller
MDA:	Metric Data Assembly
MMA:	Meteorological Monitoring Assembly
NAR:	noise adding radiometer
NASCOM:	NASA Communications
ND:	noise diode
NSS:	Network Support Subsystem
PC:	personal computer (IBM compatible)
REC:	Receiver-Exciter Controller
SETBL:	systematic error table
SPC:	Signal Processing Center
TWM:	traveling-wave maser

**Table 2. Systematic pointing error model for 34-m Az-EI antennas**

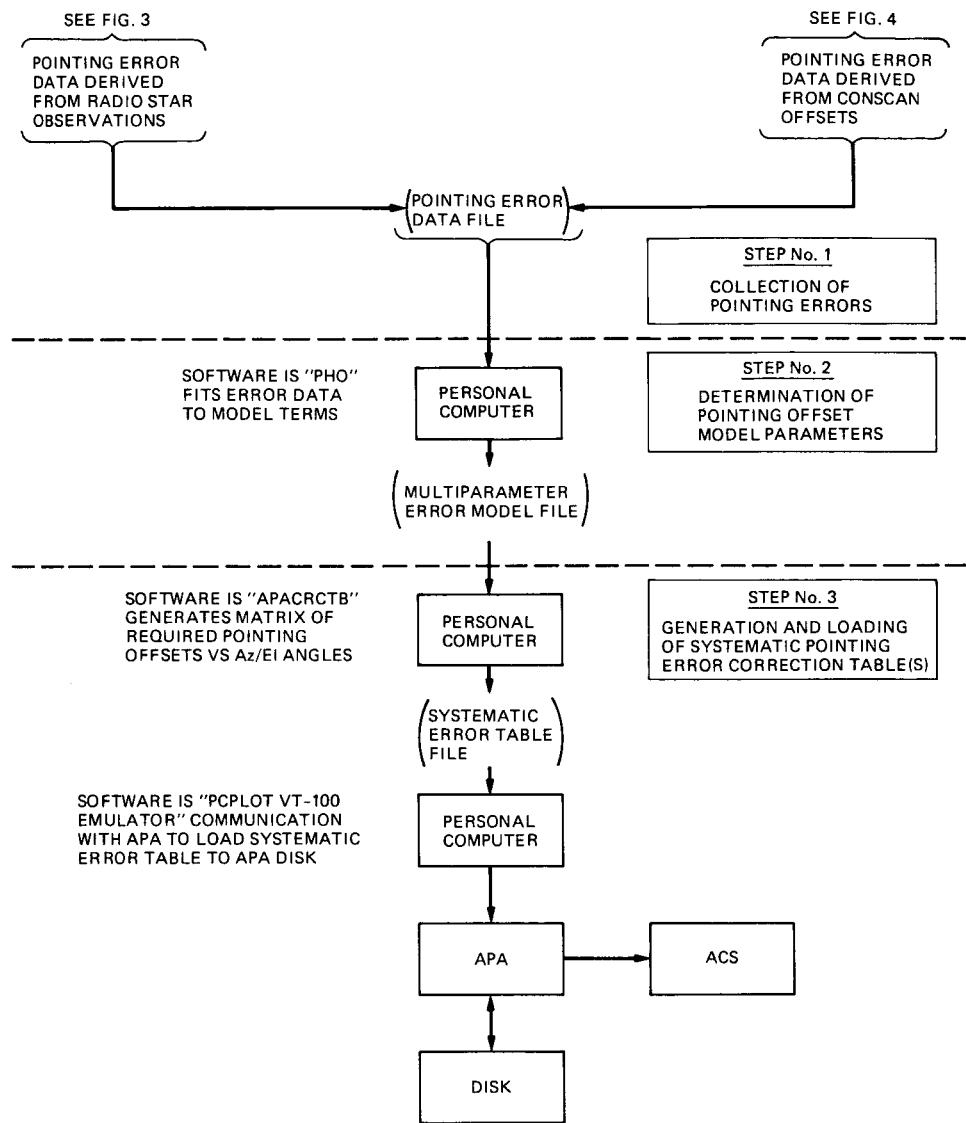
Error source	Model function for Cross-EI error	Model function for EI error
Az collimation	$P_1$	(n/a)
Az encoder bias	$P_2 \times \cos(\text{EI})$	(n/a)
Az/EI nonorthog	$P_3 \times \sin(\text{EI})$	(n/a)
Az axis tilt	$P_4 \times \sin(\text{EI}) \times \cos(\text{EI})$ $P_5 \times \sin(\text{EI}) \times \sin(\text{Az})$	$-P_4 \times \sin(\text{Az})$ $P_5 \times \cos(\text{Az})$
Source Dec	$P_6 \times \sin(\text{Az})$	$P_6 \times \sin(\text{EI}) \times \cos(\text{Az})$
EI encoder bias	(n/a)	$P_7$
Grav flex, source Dec	(n/a)	$P_8 \times \cos(\text{EI})$
Refraction	(n/a)	$P_9 \times \cot(\text{EI})$

**Table 3. Summary of results from DSS 15 Voyager 2 X-band single-axis offset Conscan snap-on tests**

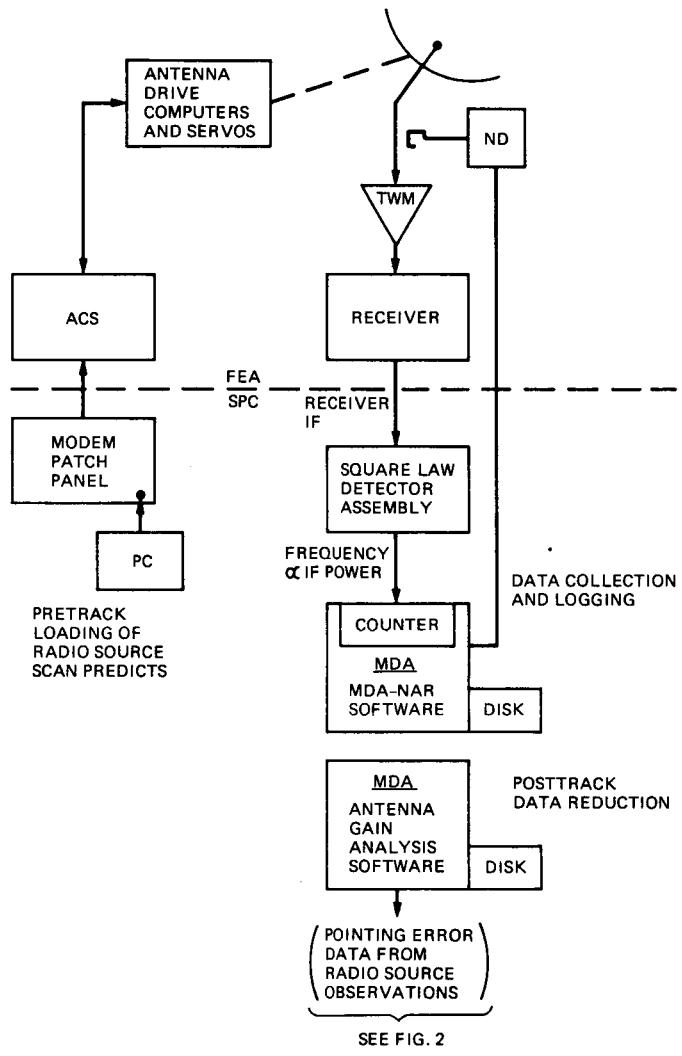
Offset change, deg	90% settling time, min	First-scan undershoot, %	Fourth-scan overshoot, %	Cross coupling, %
EI +0.020	7	40	0	18
EI -0.020	8	50	0	21
Az +0.020	4	25	5	7
Az -0.020	3	45	6	10



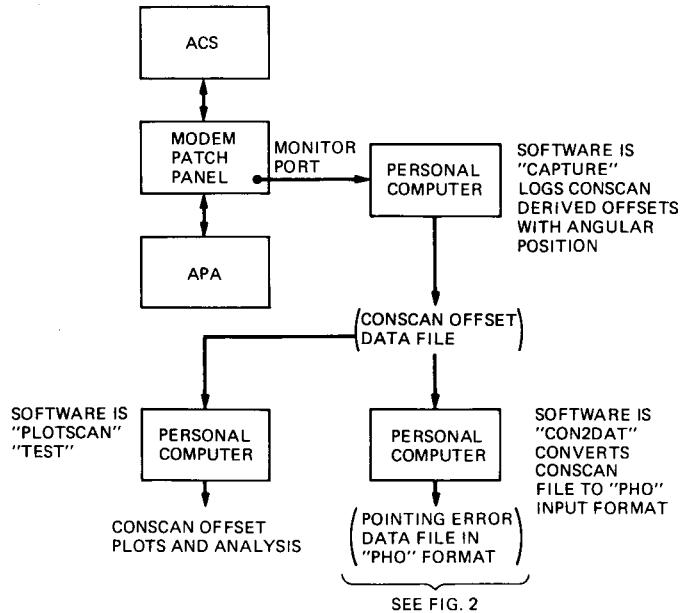
**Fig. 1. MKIVA end-to-end antenna pointing system**



**Fig. 2. Process for calibration of antenna systematic pointing errors**



**Fig. 3. Block diagram of A. Rius/MDSCC radio star antenna pointing data acquisition and analysis process**



**Fig. 4. Block diagram of S/C tracking Conscan offset data acquisition and analysis process**

DSS-14 VGR-2 DOY 026 SETBL=14EX1 Light overcast, mild, calm						
Az	EI	$\Delta Az$	$\Delta EI$	GMT	DATE	SETBL
" 126.320	008.479	+000.0007	-00.0011	12:45:51	01-26-1986	14EX1"
" 126.649	008.806	+000.0008	-00.0012	12:48:02	01-26-1986	14EX1"
" 126.978	009.131	+000.0011	-00.0012	12:49:51	01-26-1986	14EX1"
" 127.313	009.453	+000.0015	-00.0012	12:51:51	01-26-1986	14EX1"
" 206.930	026.401	+000.0036	-00.0006	18:30:19	01-26-1986	14EX1"
" 207.401	026.216	+000.0031	-00.0003	18:32:30	01-26-1986	14EX1"
" 207.869	026.025	+000.0031	-00.0006	18:34:19	01-26-1986	14EX1"
" 208.334	025.832	+000.0030	-00.0006	18:36:19	01-26-1986	14EX1"
" 208.797	025.636	+000.0027	-00.0006	18:38:20	01-26-1986	14EX1"

Az, EI,  $\Delta Az$ ,  $\Delta EI$  in degrees

SETBL=Identification of Systematic Error Correction Table in ACS

**Fig. 5. Sample of output from R. Livermore/CDSCC Conscan offset data acquisition program**

**Fig. 6. Plots and summary analyses of Conscan offset data from Voyager 2 tracking by calibrated MKIVA DSN antennas**

**Discussion:** The plots and analysis summaries are reproductions of outputs from PLOT.BAS and CSN-ANAL.EXE programs using data from CAPTURE.BAS (see text, Subsection V.A.2, and the Appendix). These products can be made at the complexes right after a tracking pass, and the data for making them can be sent immediately to JPL by electronic mail.

The headings of the plots and analysis summaries record support parameters such as V214D021 (meaning Voyager 2; DSS 14; Day of Year 21).

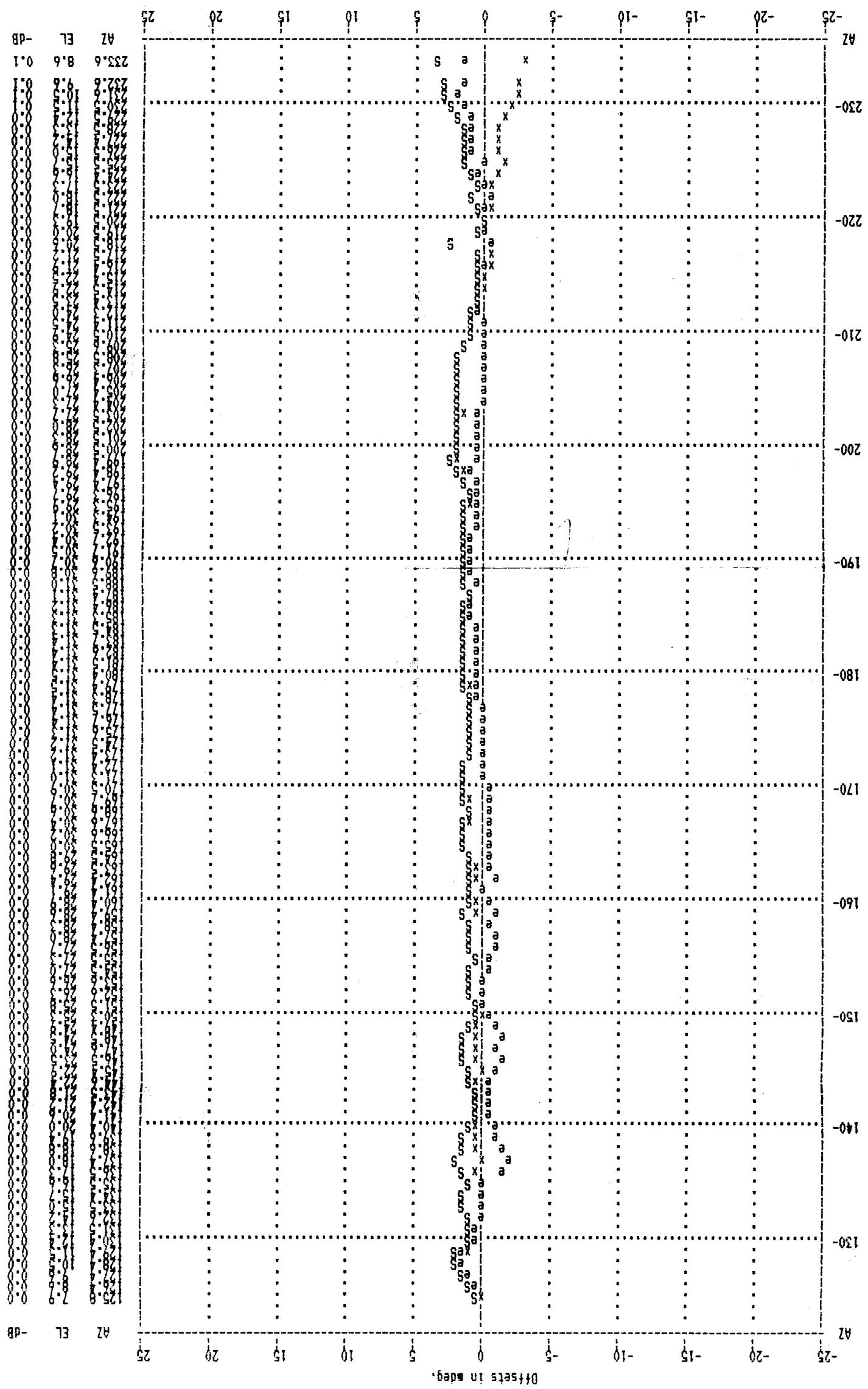
The plots are Conscan offsets vs Az (or Ha for Ha-Dec antennas). Az and El angles for each offset point are in the right columns. Also, the dB losses (-dB) calculated to result from each offset point are in the rightmost column. On the plots, elevation offsets are "e," the cross-elevation offsets are "x," and the space (or total beam) offsets are "S." The "S" are calculated as the square root of the sum of the squares of "e" and "x;" the positive value is presented. The "S" are always plotted; for overlaps of "e" and "x," the "e" survives.

The analysis summaries are statistical calculations on the file data. They provide a succinct view of the antenna pointing performance: in particular, mean space offset and mean beam pointing loss that would be incurred under blind pointing. Other analysis results are available from the program, but are not shown here, e.g., specific occurrences of offsets where the pointing loss would exceed the specified maximum.

The data represented in the plots and analysis summaries are from the antennas in their prime operating modes for Voyager 2 UE support: all predict driven and with precision (Master Equatorial) control on the 64-m antennas. There is one exception. Figure 6(a2) shows results from an unusual DSS 14 pass during which the first part of the pass was in the prime precision mode and the second part was in the backup computer control mode. The periods of use of the two modes are indicated on the plot; a separate analysis summary for each mode is shown. The difference in tracking quality of the two modes is illustrated.

For reference in reviewing the plots and analyses, the pointing specification adopted was:  $\leq 6$  mdeg total beam pointing error ( $\leq 4$  mdeg each for equal El and cross-El errors) for the 64-m antennas;  $\leq 11$  mdeg total beam pointing error ( $\leq 8$  mdeg each for equal El and cross-El errors) for the 34-m antennas. That specification limits the allowable pointing loss to 0.3 dB, maximum.

ORIGINAL PAGE IS  
OF POOR QUALITY



RECEIVING PAGE BLANK NOT FILLED

Fig. 6(a1). DSS 14 64-m precision mode (GDSCC)

FOLDOUT FRAME

FOLDOUT FRAME

FOLDOUT FRAME

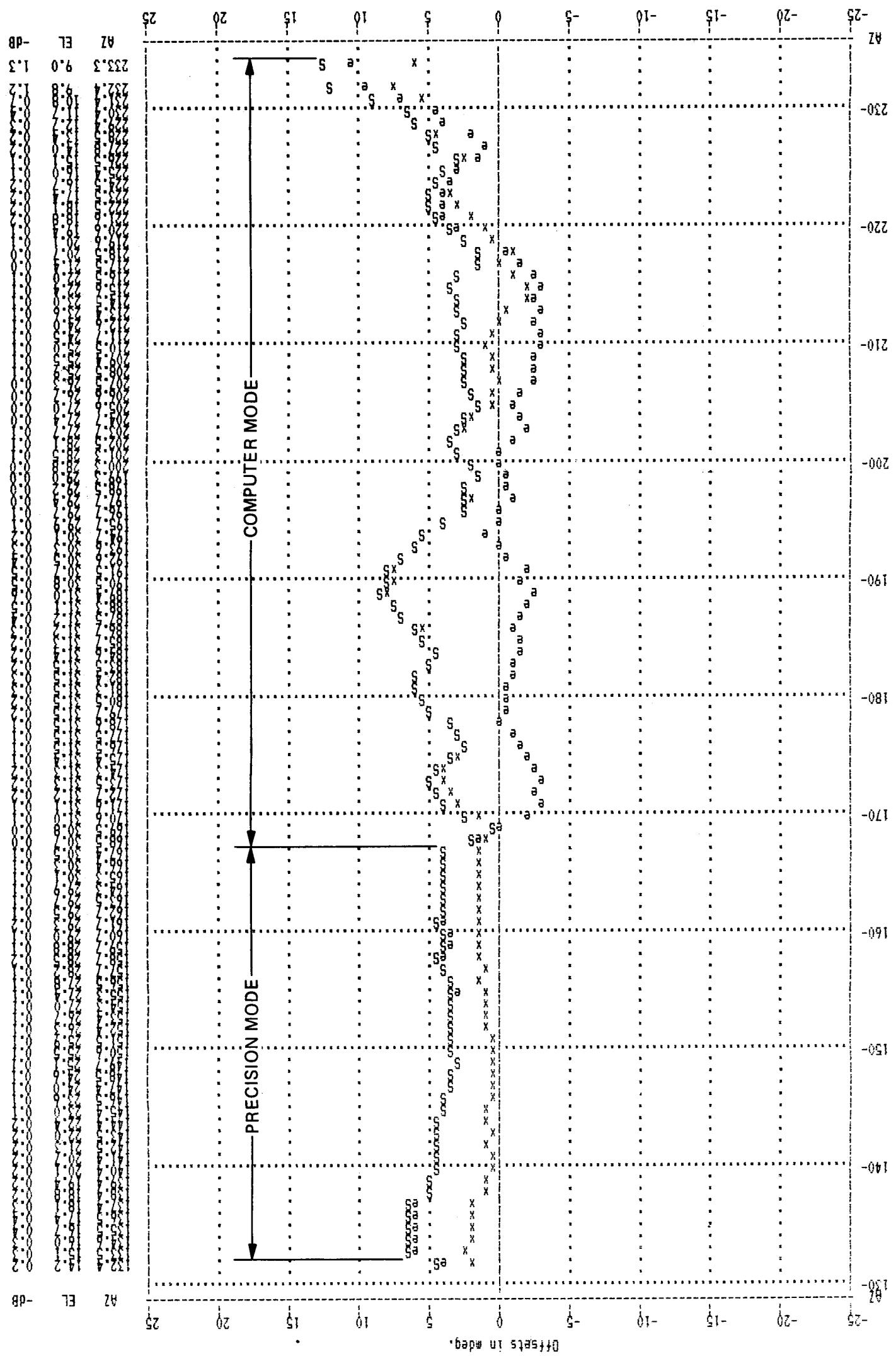


Fig. 6(a2). DSS 14 64-m precision mode and computer control mode (GDSCC)

FOLDOUT FRAME

FOLDOUT FRAME

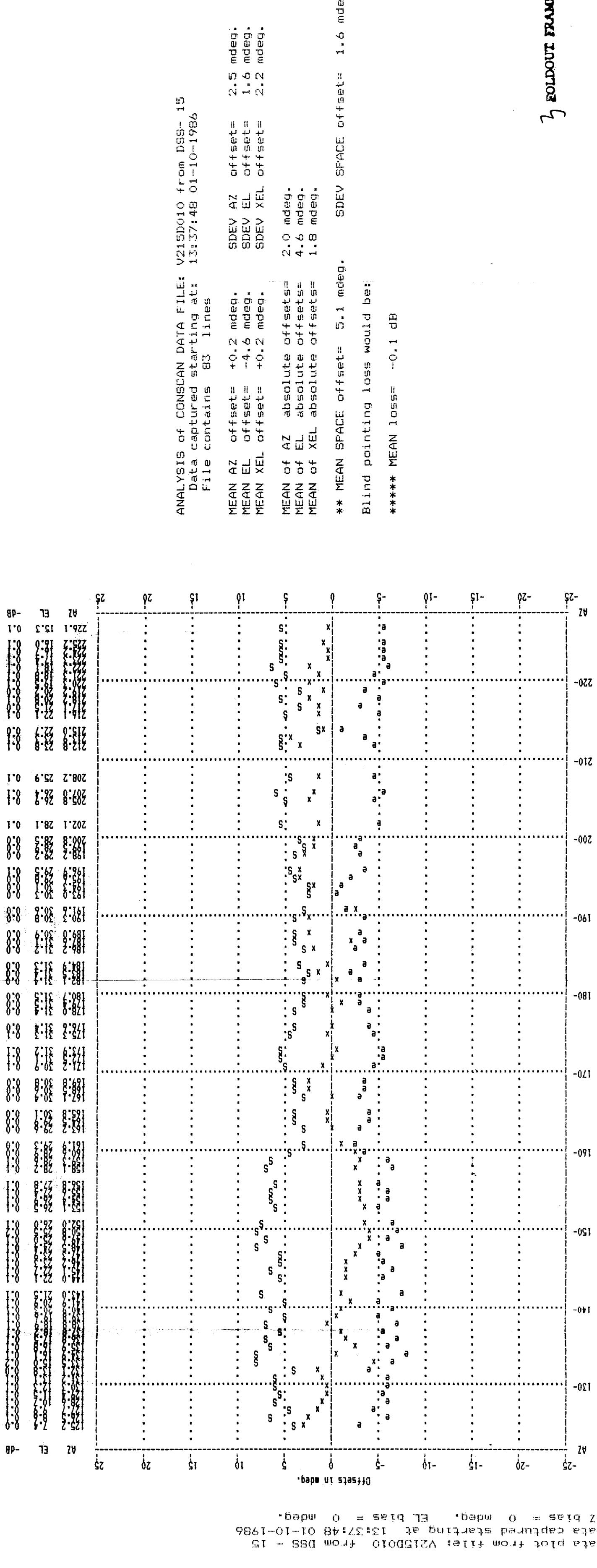
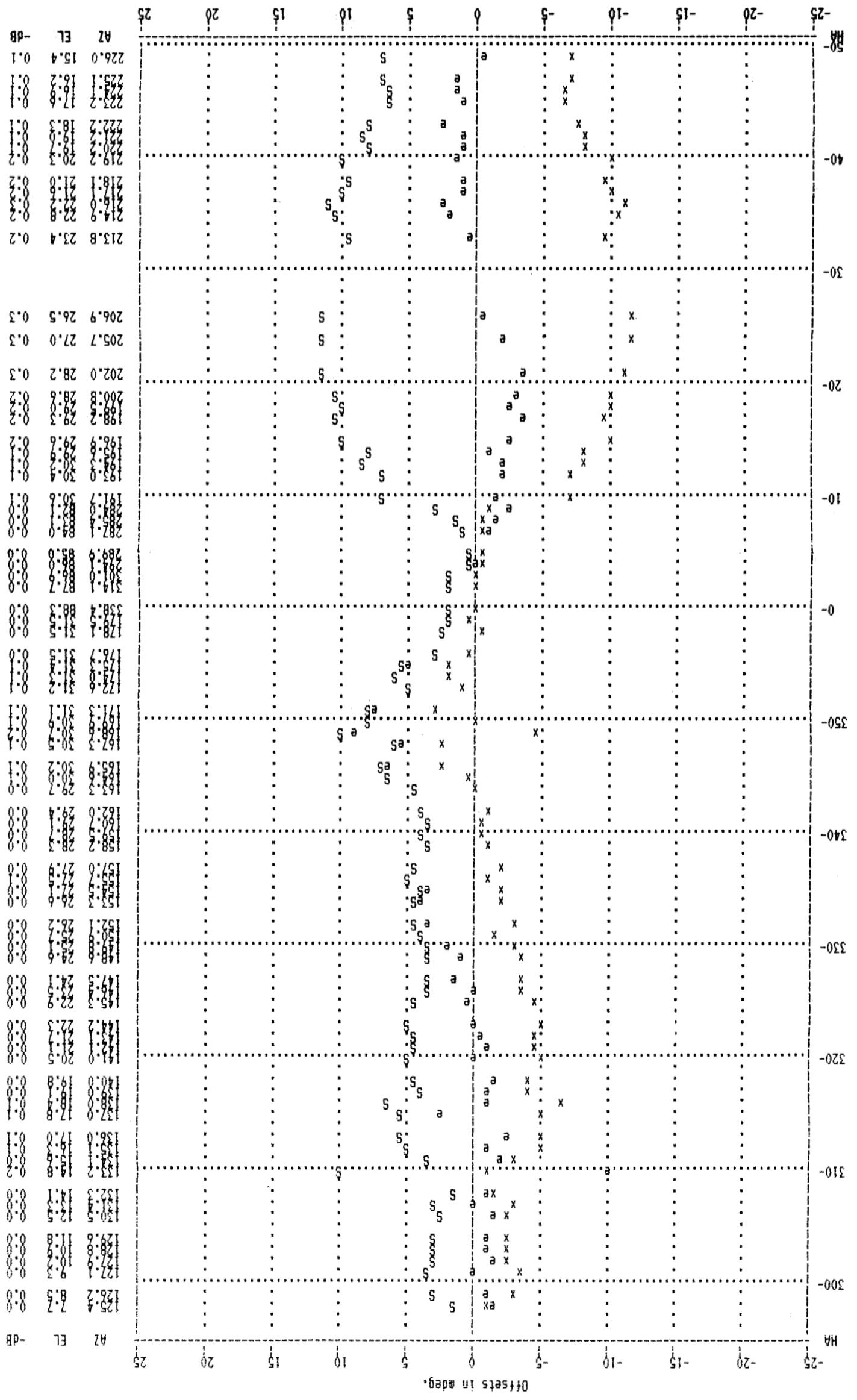


Fig. 6(b). DSS 15 34-m AZ-EI (GDSC)



Data plot from file: V212D010 from DSS - 12  
 Data captured starting at 13:37:53 01-10-1986  
 AZ bias = 0 mdeg. EL bias = 0 mdeg.

Fig. 6(c). DSS 12 34-m Ha-Dce (GDSCC)

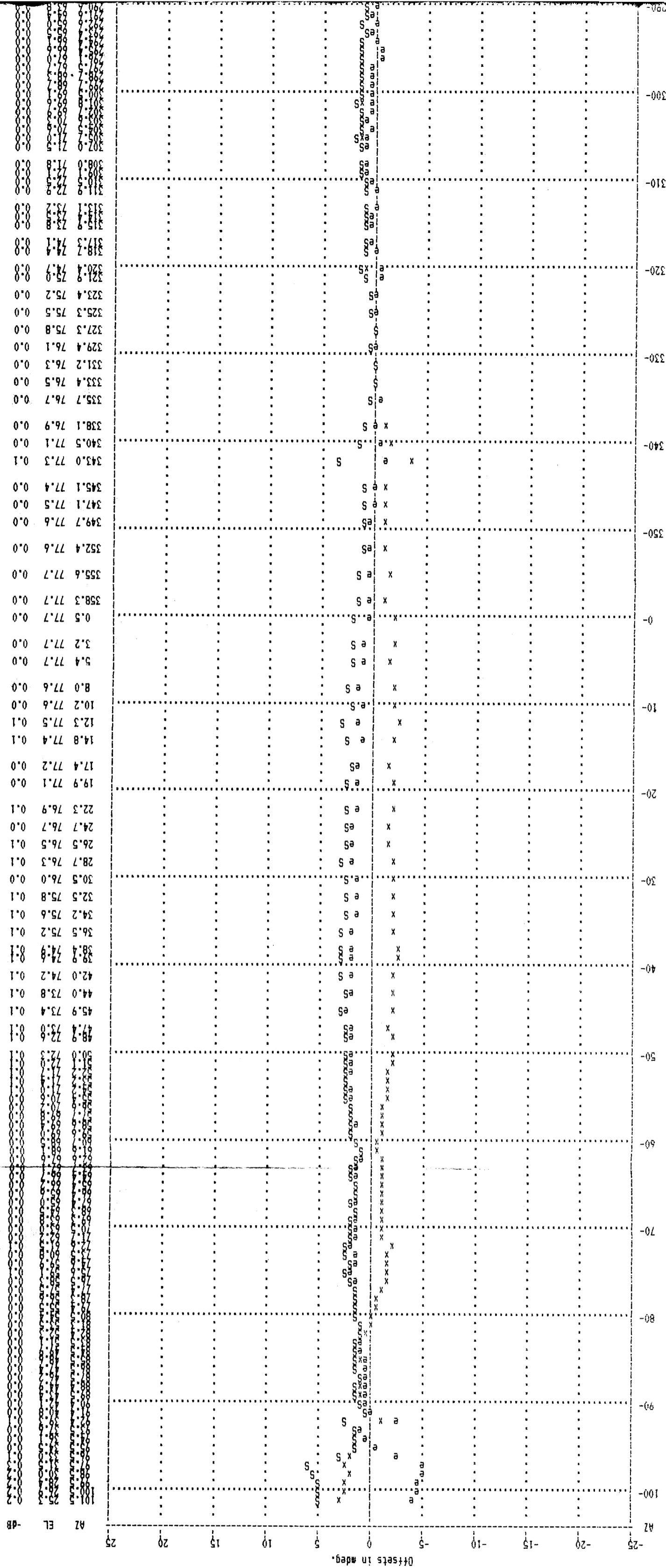
FOLDOUT FRAME

FOLDOUT FRAME

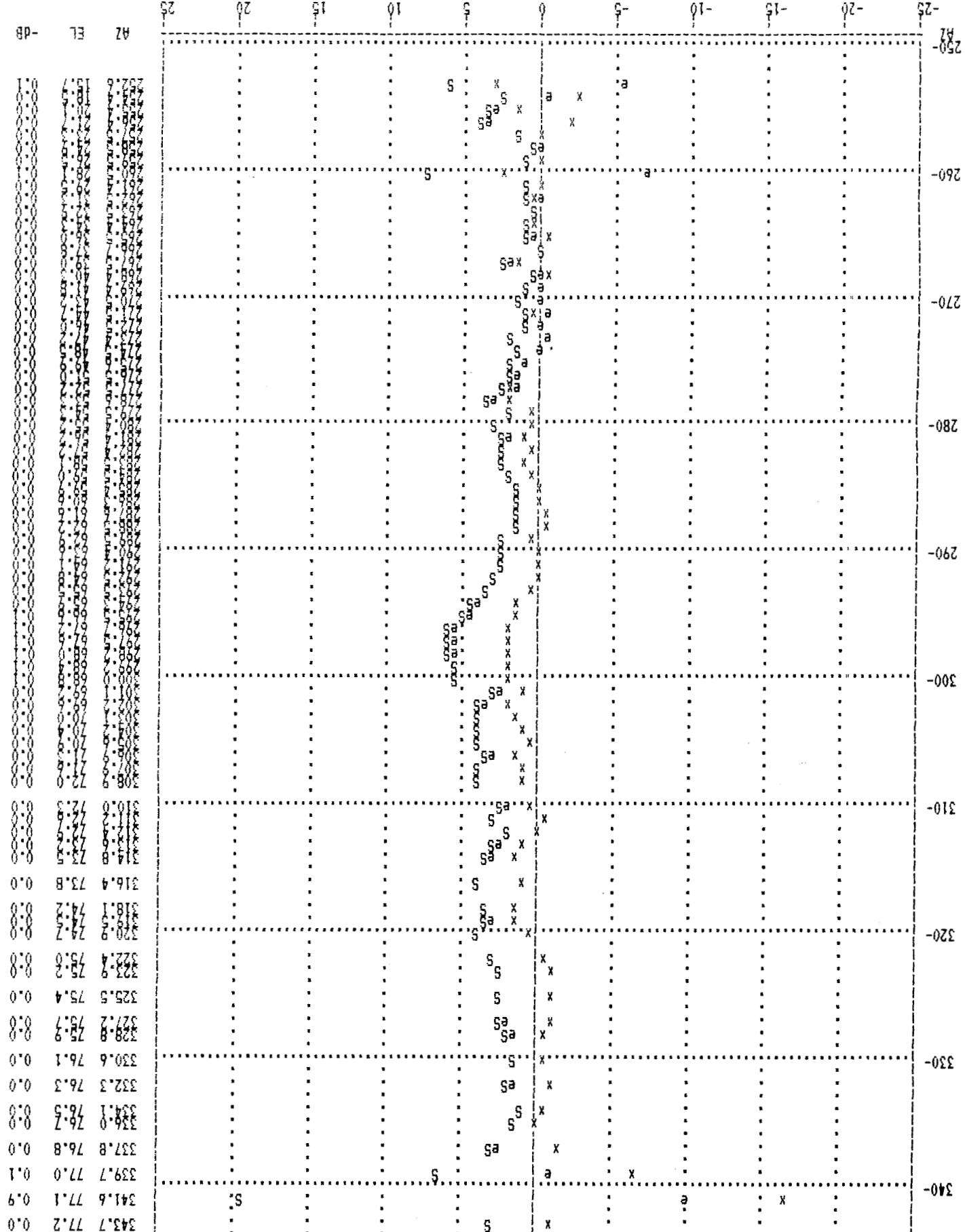
ORIGINAL PAGE IS  
OF POOR QUALITY

FOLDOUT ERASE

FOLDOUT FRAME



ORIGINATE POINTS ID  
OF PICTURE



ANALYSIS of CUNSCAN DATA FILE: V2450018.F from DSS-- 45  
Data captured starting at: 19:18:49 01-18-1986  
File contains 275 lines

MEAN AZ offset= -0.0 mdeg. SDEV AZ offset= 6.3 mdeg.  
MEAN EL offset= +0.6 mdeg. SDEV EL offset= 3.2 mdeg.  
MEAN XEL offset= +0.3 mdeg. SDEV XEL offset= 3.1 mdeg.

MEAN of AZ absolute offsets= 3.0 mdeg.  
MEAN of EL absolute offsets= 2.0 mdeg.  
MEAN of XEL absolute offsets= 1.6 mdeg.

\*\* MEAN SPACE offset= 2.8 mdeg. SDEV SPACE offset= 3.5 mdeg.

Blind pointing loss would be:

\*\*\*\*\* MEAN LOSS= -0.0 dB

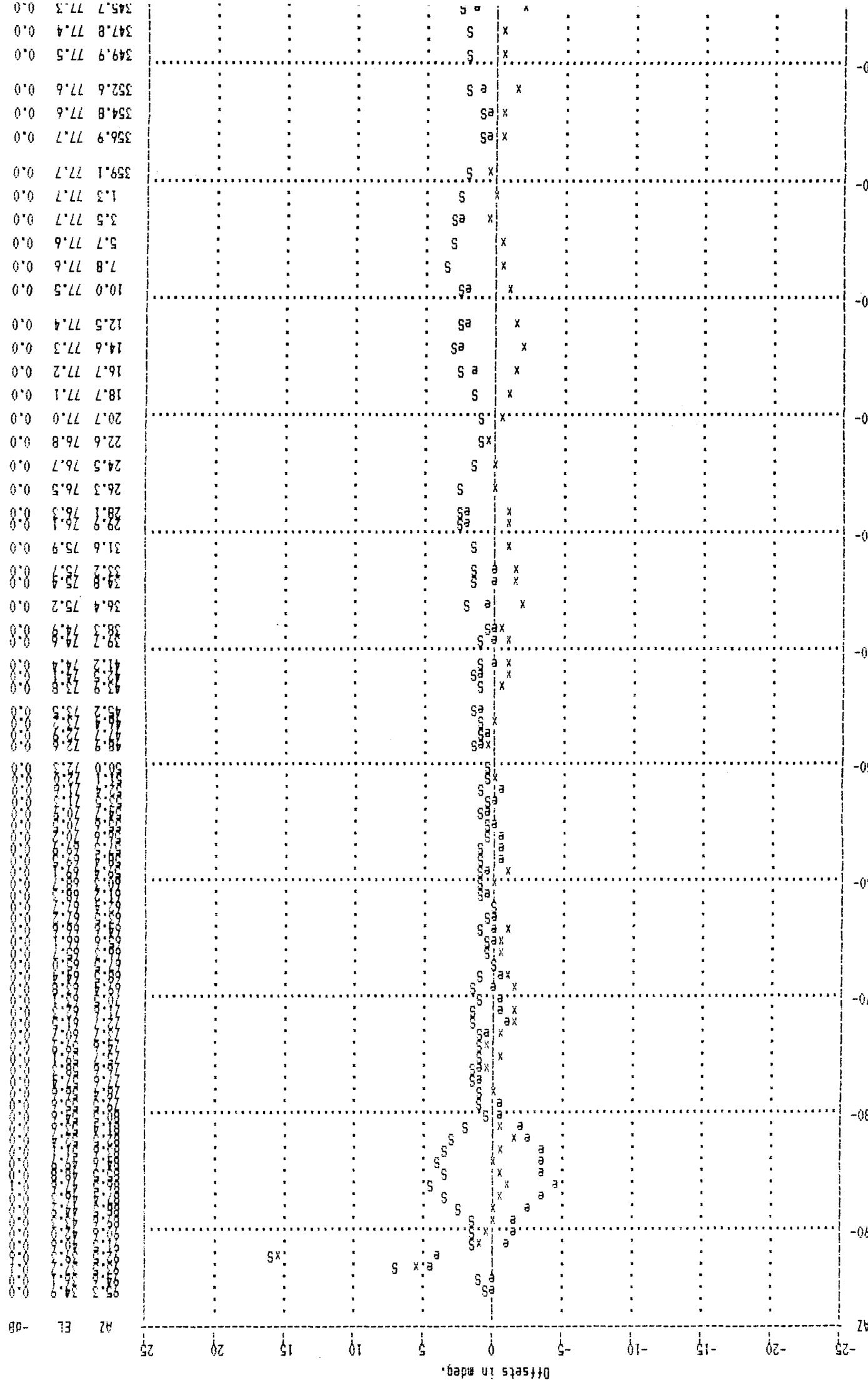
7. HOLDOUT FRAME

4. HOLDOUT FRAME

### DROUDOUT FRAME

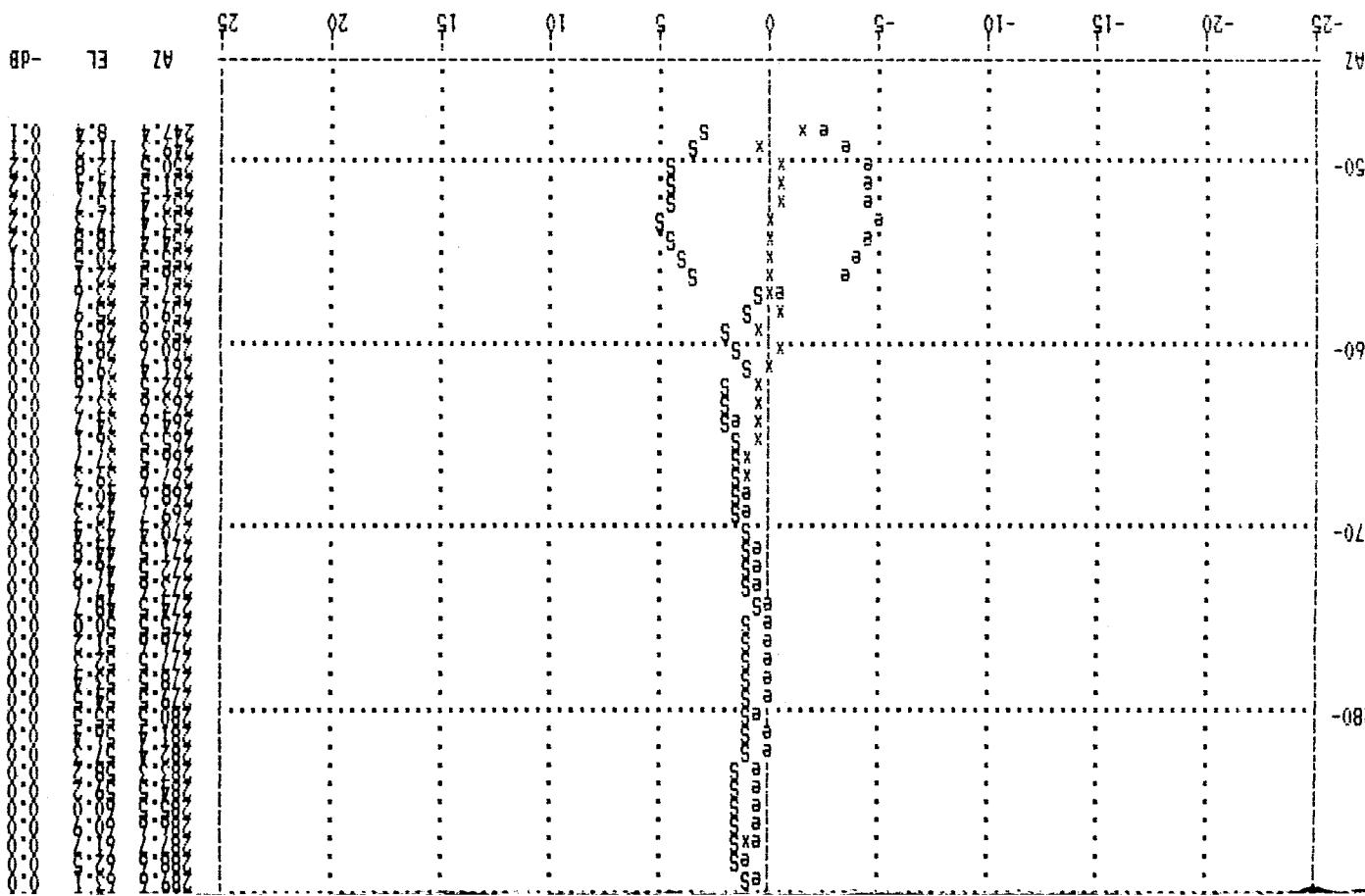
### DROUDOUT FRAME

ORIGINAL PAGE IS  
OF POOR QUALITY.



Data plot from file: V245D01B.F from DSS - 45  
Data captured starting at 19:18:49 01-18-1986  
AZ bias = 0 deg. EL bias = 0 deg.

Fig. 6(e). DSS 45 34-m AZ-EI (CDSCC)



ANALYSIS of CONSCAN DATA FILE: V243D018.P1F from DSS- 43  
Data captured starting at: 16:28:43 01-18-1986  
File contains 300 lines

MEAN AZ offset= -0.1 mdeg.  
SDEV AZ offset= 3.3 mdeg.  
MEAN EL offset= -0.0 mdeg.  
SDEV EL offset= 2.1 mdeg.  
MEAN XEL offset= +0.3 mdeg.  
SDEV XEL offset= 1.3 mdeg.

MEAN of AZ absolute offset= 2.5 mdeg.  
MEAN of EL absolute offset= 1.5 mdeg.  
MEAN of XEL absolute offset= 1.1 mdeg.

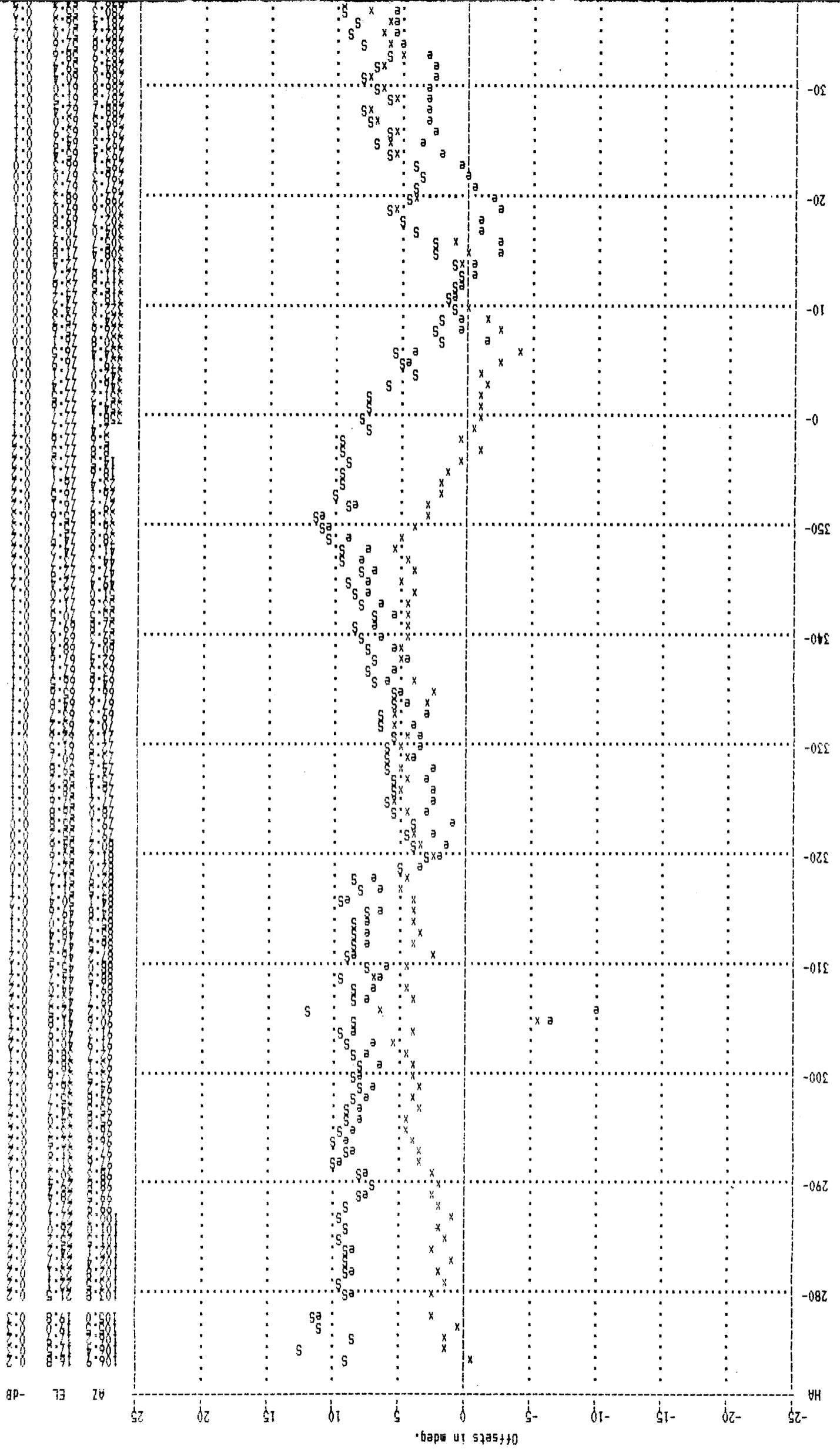
\*\* MEAN SPACE offset= 2.1 mdeg.  
SDEV SPACE offset= 1.3 mdeg.

Blind pointing loss would be:  
\*\*\*\*\* MEAN loss= -0.0 dB

4 FOLDOUT FRAMES

Fig. 6(d). DSS 43 64-m precision mode (CDSSCC)

ORIGINAL PAGE IS  
OF POOR QUALITY



Data plot from file: V242D020.A from DSS - 42  
Data captured starting at 17:40:20 01-20-1986  
AZ bias = 0 deg. EL bias = 0 deg.

2 FOLDOUT FRAME

1 FOLDOUT FRAME

ANALYSIS of CONSCAN DATA FILE: V263D021 from DSS- 63  
Data captured starting at: 06:49:17 01-21-1986  
File contains 220 lines

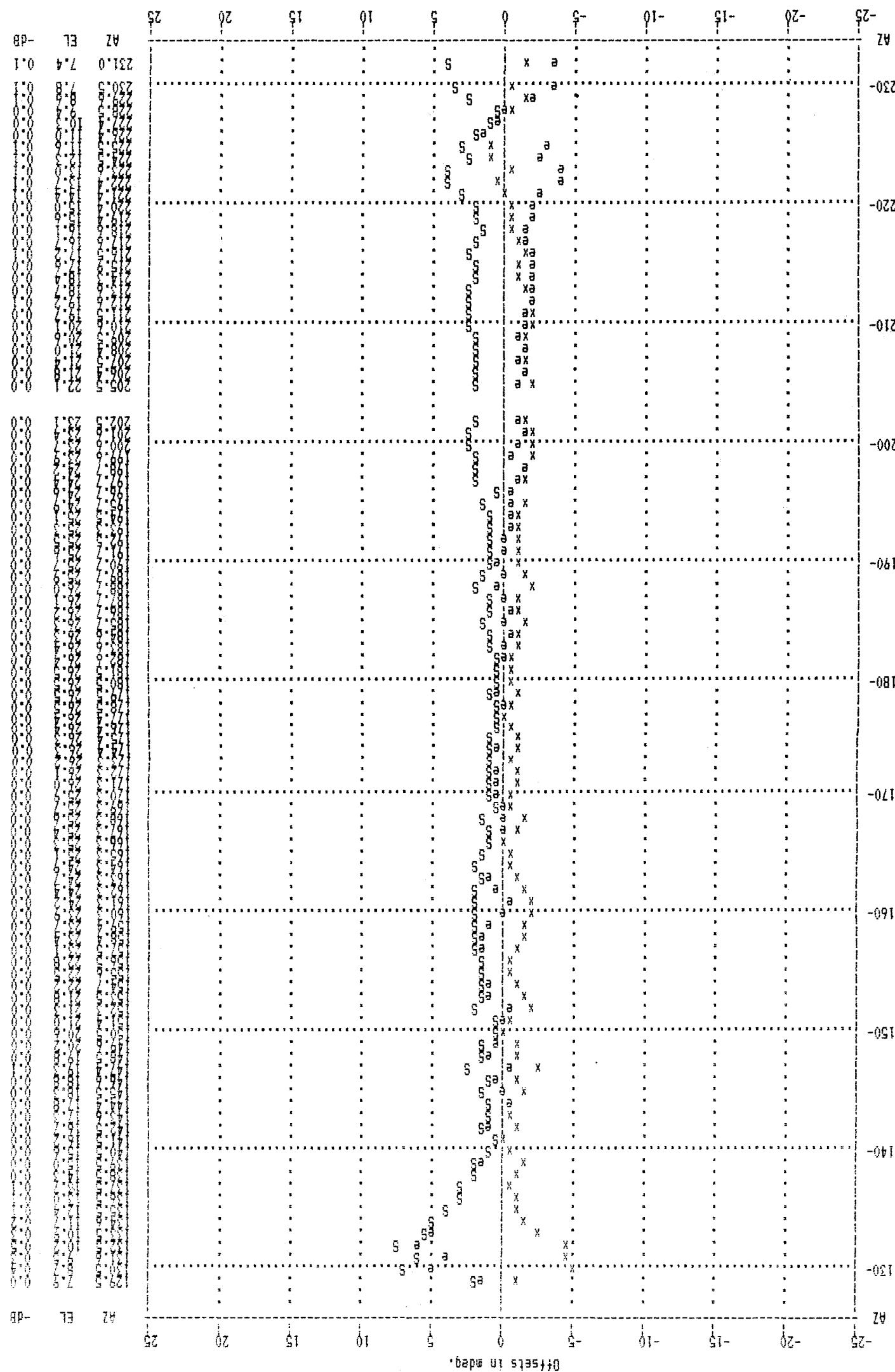
MEAN AZ offset= -1.2 mdeg. SDEV AZ offset= 1.1 mdeg.  
MEAN EL offset= +0.1 mdeg. SDEV EL offset= 2.0 mdeg.  
MEAN XEL offset= -1.1 mdeg. SDEV XEL offset= 1.1 mdeg.  
  
MEAN of AZ absolute offset= 1.3 mdeg.  
MEAN of EL absolute offset= 1.4 mdeg.  
MEAN of XEL absolute offset= 1.2 mdeg.  
  
\*\* MEAN SPACE offset= 2.0 mdeg. SDEV SPACE offset= 1.4 mdeg.

Blind pointing loss would be:

\*\*\*\* MEAN loss= -0.0 dB

3 FOLDOUT FRAME

ORIGINAL PAGE IS  
OF POOR QUALITY



**Fig. 6(g).** DSS 63 64-m precision mode (MDSSC)

Data plot from file: V262D021 from DSS - 63  
Data captured starting at 06:49:17 01-21-1986  
AZ bias = 0 deg. EL bias = 0 deg.

228

FOLDOUT FRAME

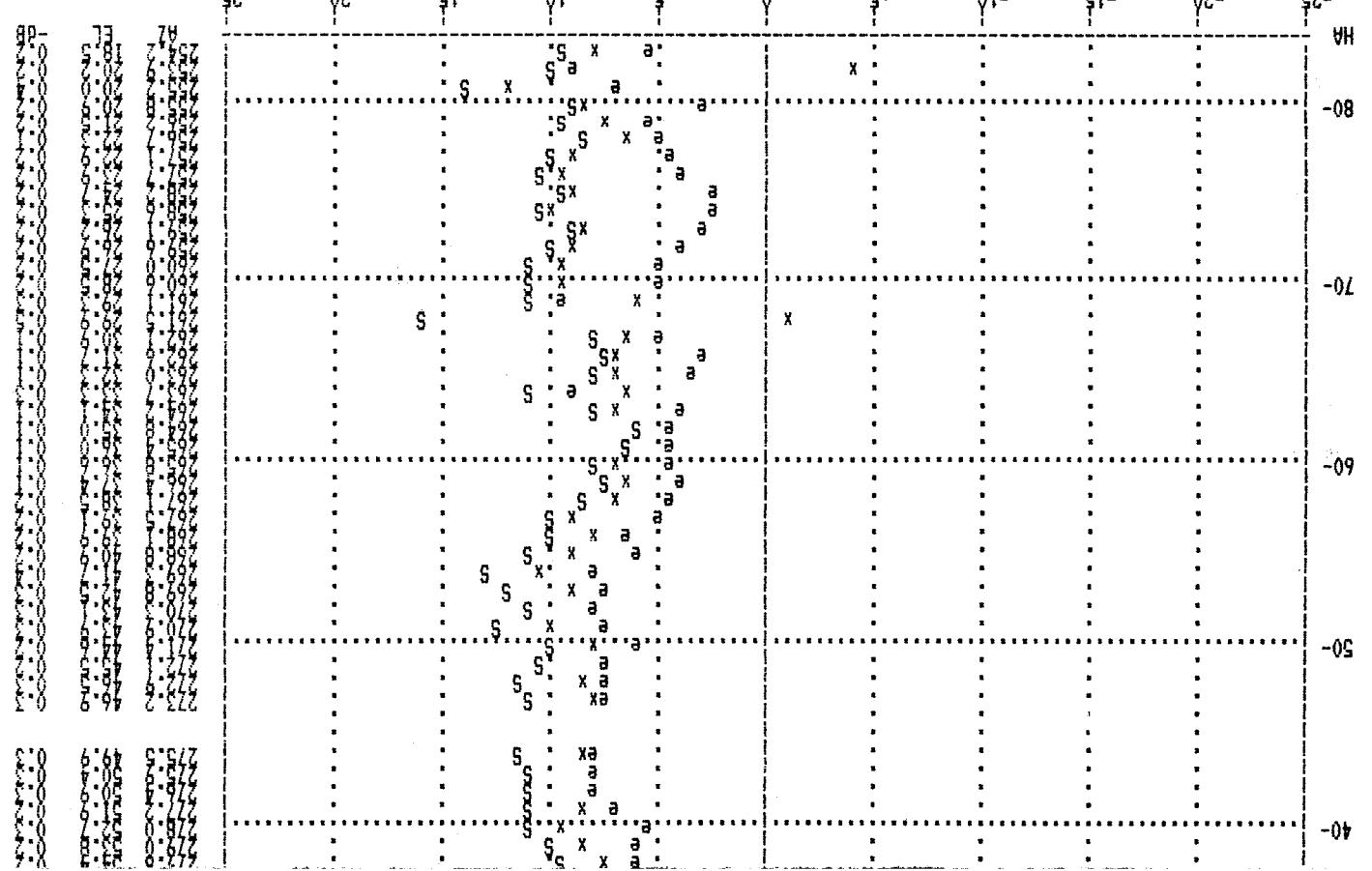
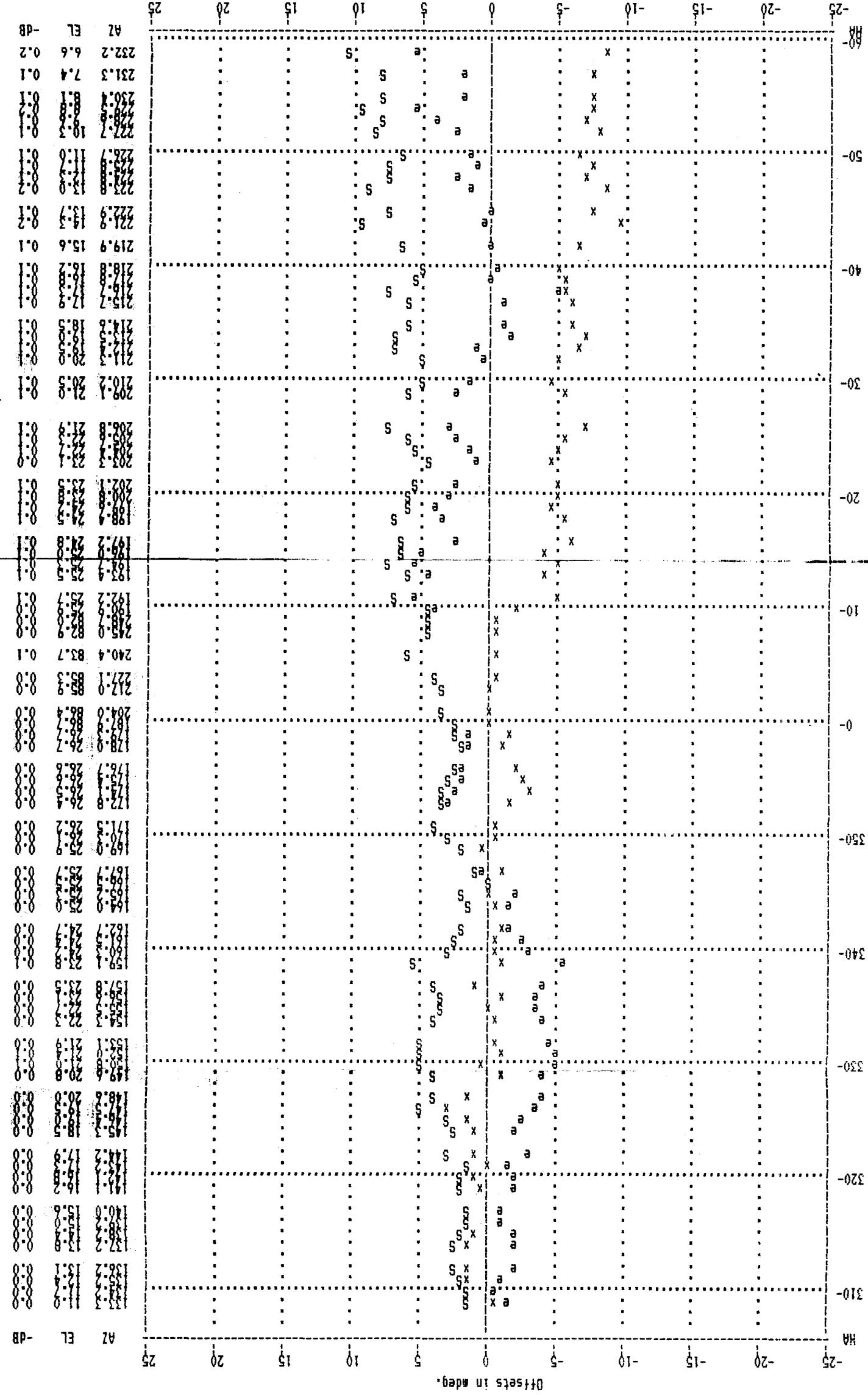


Fig. 6(f). DSS 42 34-m Ha-Dec (CDSCC)

ORIGINAL PAGE IS  
OF POOR QUALITY

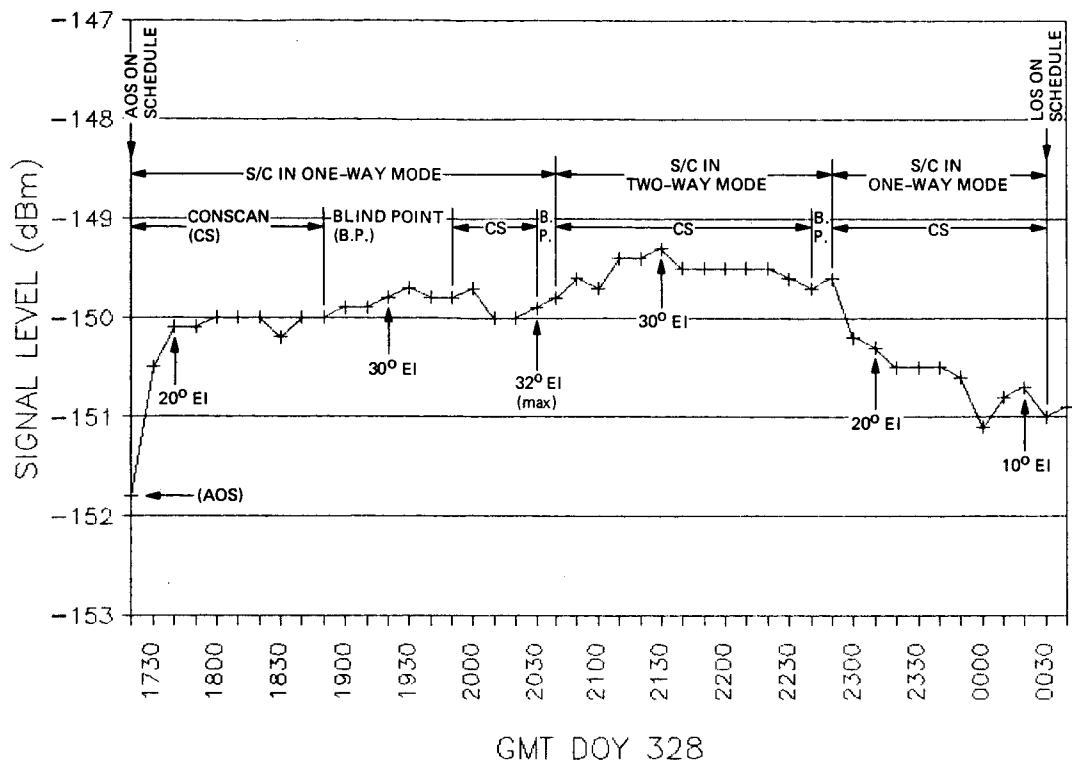


Data Plot from file: V261D338 from DSS - 61  
Data captured starting at 08:45:00 12-04-1985  
AZ bias = 0 mdeg. EL bias = 0 mdeg.

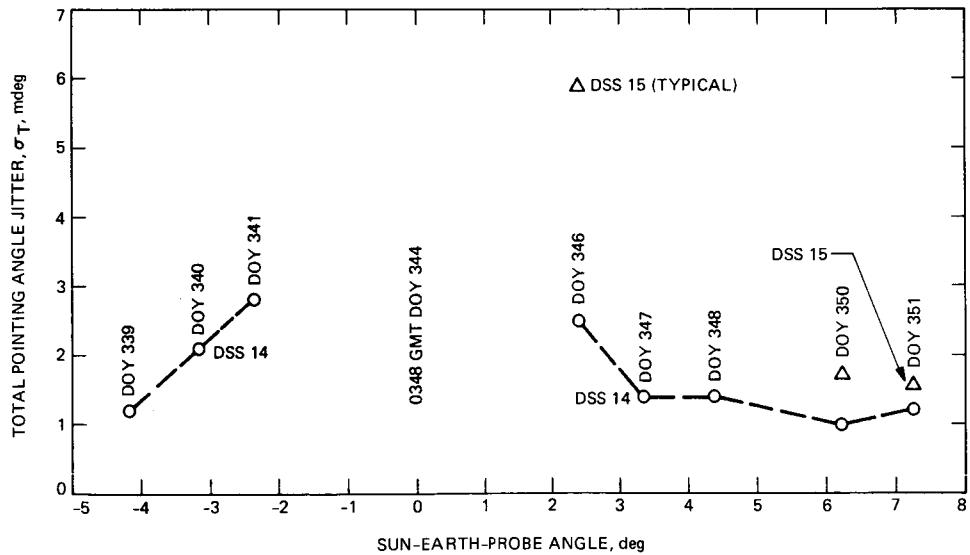
Fig. 6(h). DSS 61 34-m Ha-Dec (MDSCC)

FOLDOUT FRAME

2 FOLDOUT FRAME



**Fig. 7. Received signal level during Voyager 2 tracking in Conscan and blind pointing modes at DSS 14**



**Fig. 8. Pointing angle jitter during Voyager 2 X-band tracking near solar conjunction, DSS 14 and DSS 15**

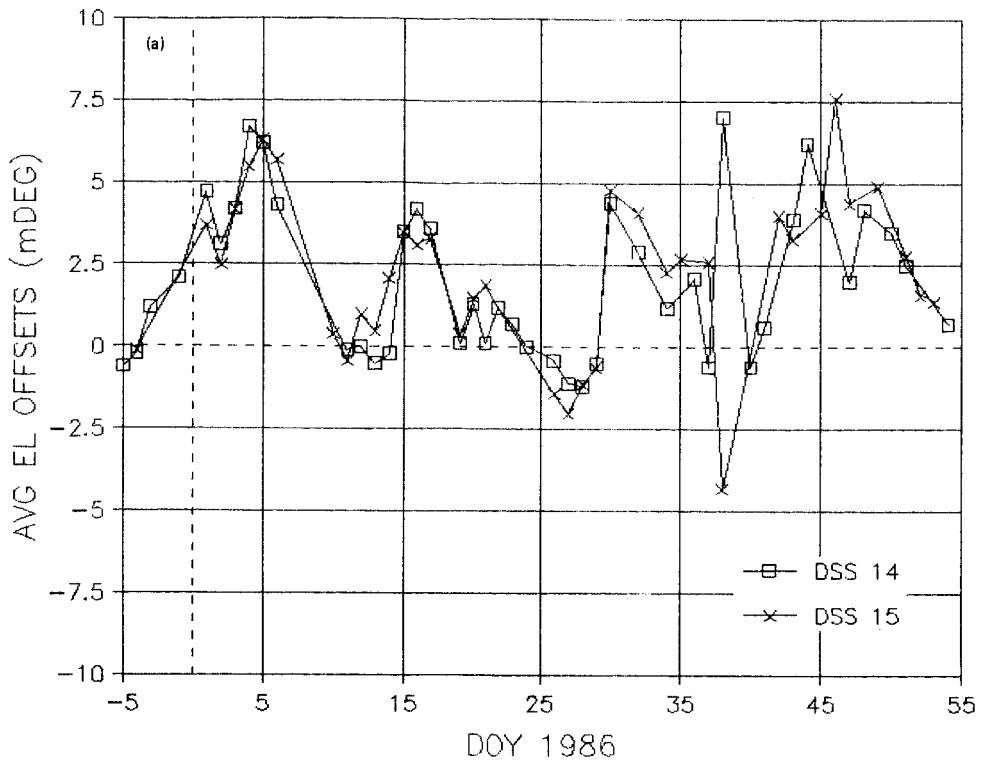
**Discussion:** The total pointing angle jitter,  $\sigma_T$ , is calculated as the square root of the sum of the squares of the elevation axis and the cross-elevation axes jitters. The jitter of each axis was estimated from examination of the Conscan offset plots for elevation angles above 20 degrees.

**Fig. 9. Comparison of pass average Conscan offsets from Voyager 2 X-band tracking at DSS 14 and DSS 15**

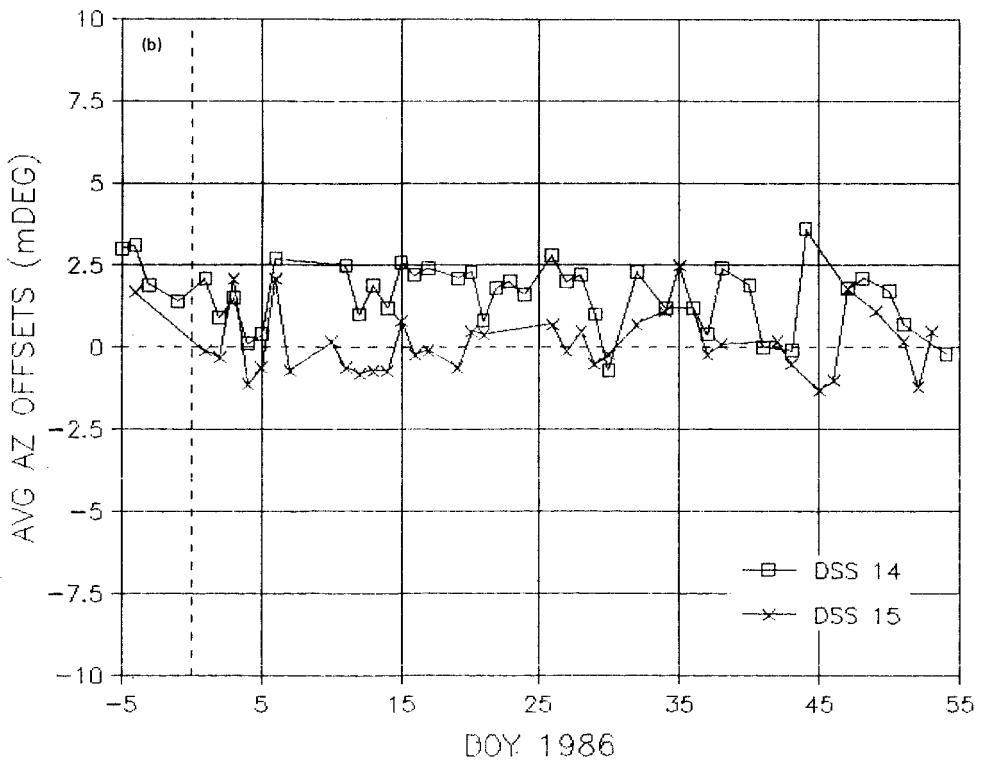
**Discussion:** On most days of the period covered, the Voyager 2 S/C was tracked by DSS 14 and DSS 15 simultaneously — the antennas were being arrayed for the near-encounter support. Uranus closest approach was on DOY 23.

Each point represents the arithmetic average of the offsets during the pass, sampled at 1-deg azimuth increments. A full-length Voyager 2 pass at GDSCC covered 125 to 235 deg azimuth and 8 to 32 deg elevation.

Generally, the elevation offsets at the two antennas show high correlation; the azimuth offsets show low correlation (see text). There are inconsistent elevation offsets on DOY 38 (Fig. 9(a)). The pointing predicts for that day were examined carefully, and exonerated. The inconsistency was not resolved.



**Fig. 9(a). Average elevation offsets**



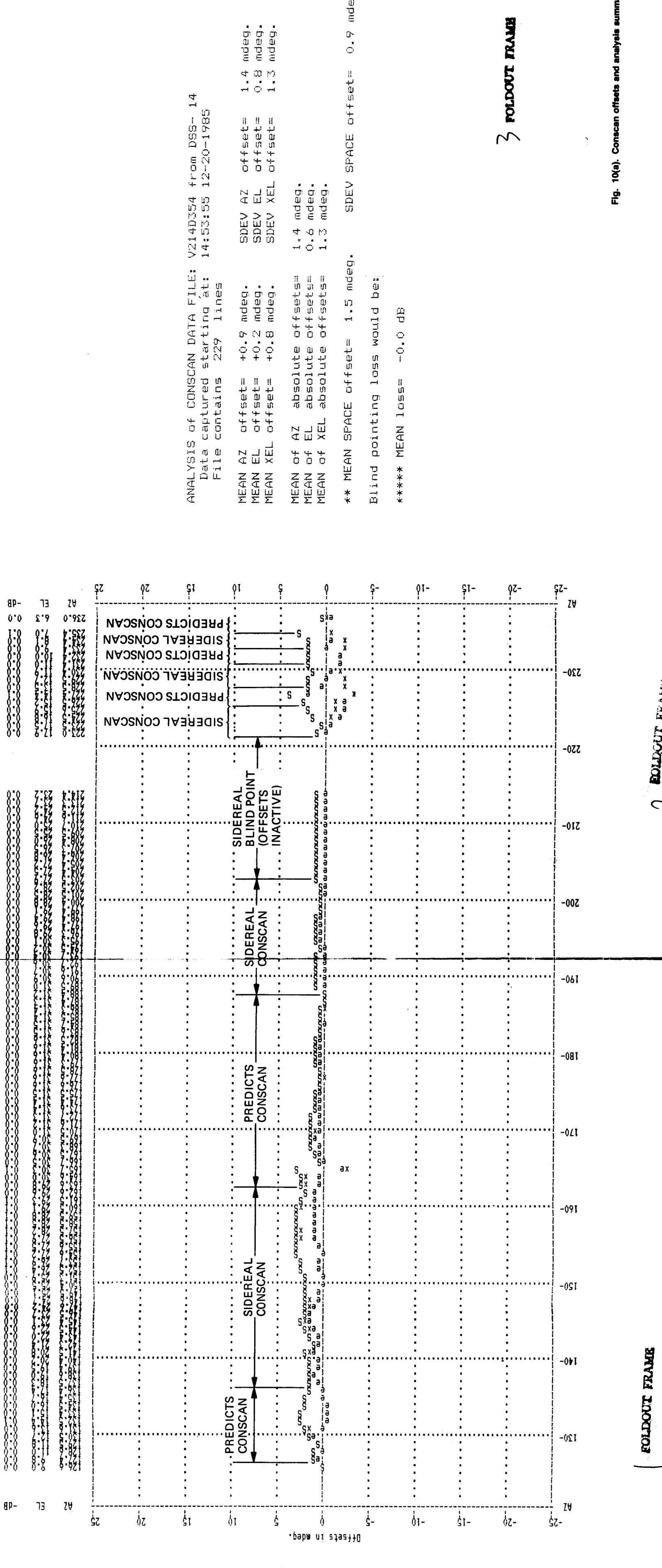
**Fig. 9(b). Average azimuth offsets**

**Fig. 10. Conscan offsets and received signal level during Voyager 2 tracking in sidereal and predict modes at DSS 14**

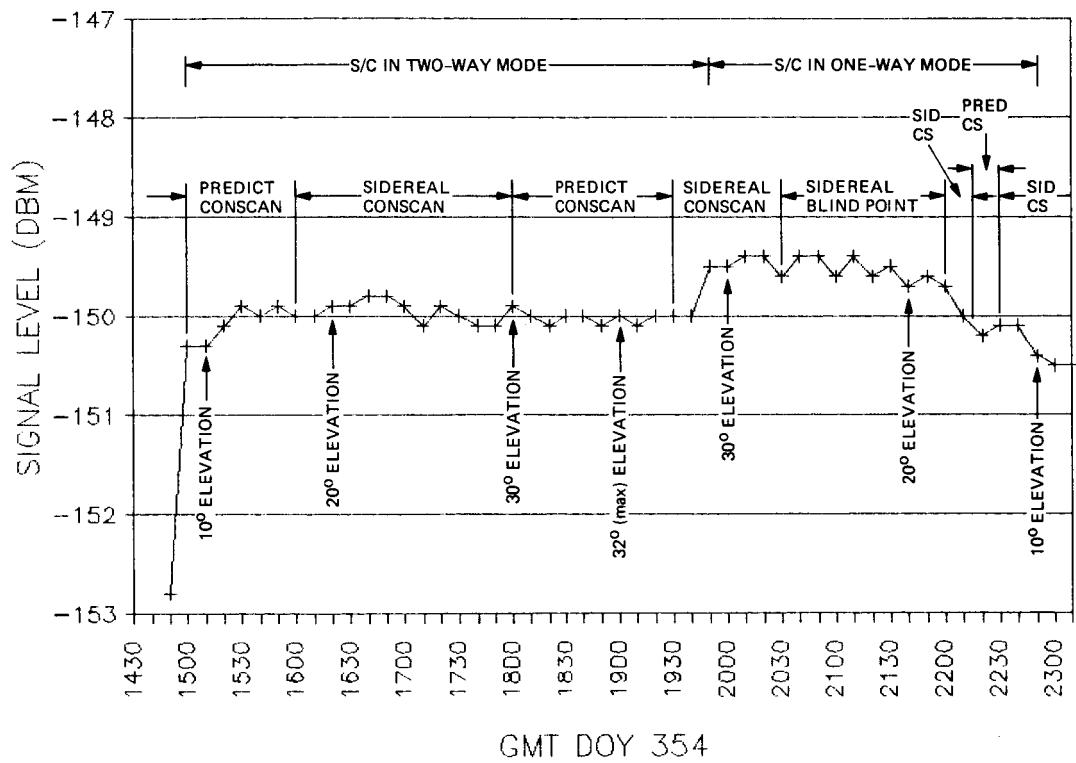
**Discussion:** The format of the Conscan offset plots and the summary analysis of Fig. 10(a) are the same as for Fig. 6, and the discussion presented there applies. The periods of operation in predict and sidereal Conscan, and sidereal blind-point modes are shown.

The received signal-level plot of Fig. 10(b) shows the periods of use of the different modes, and the antenna elevation angle at intervals during the track.

ORIGINAL PAGE IS  
OF POOR QUALITY

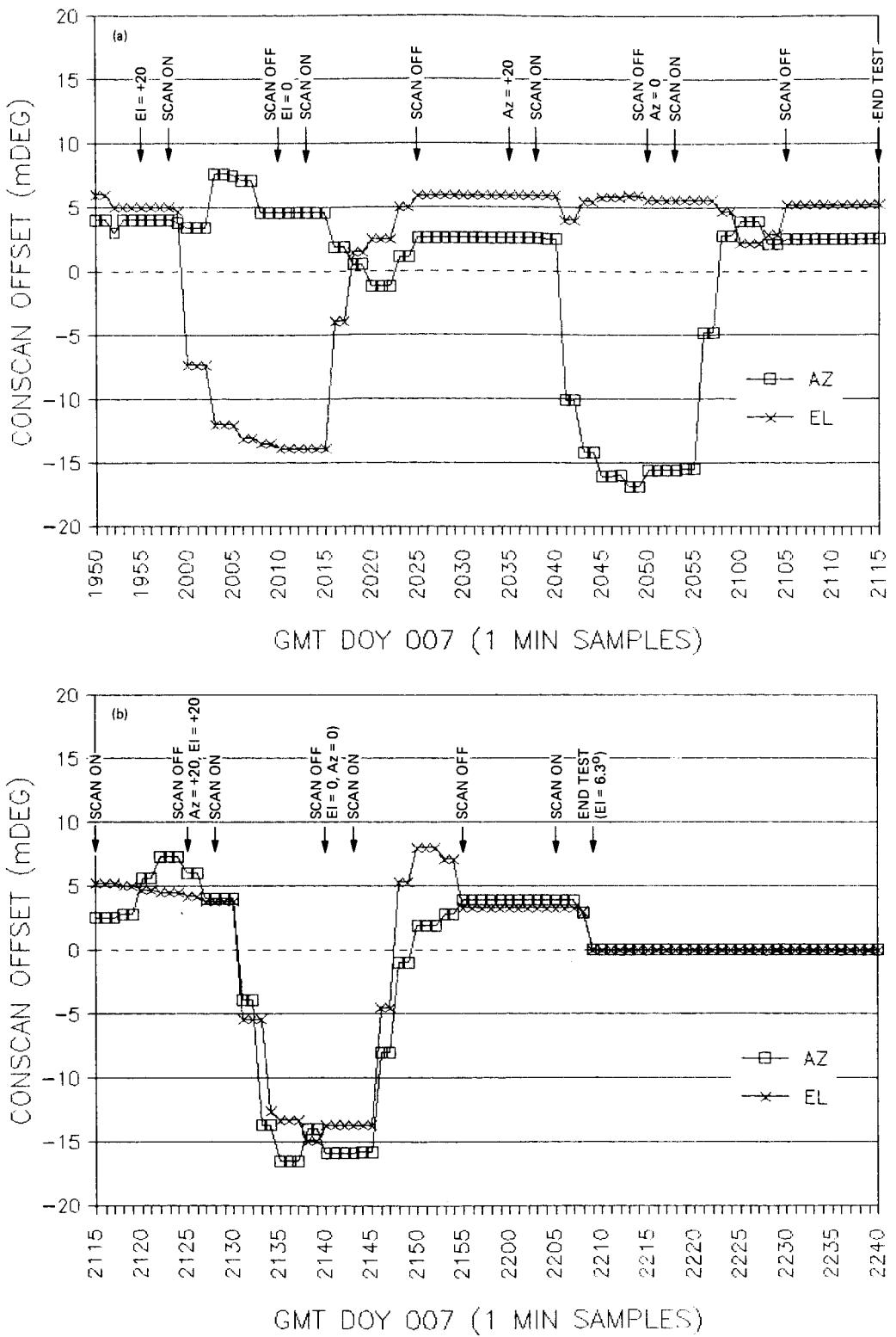


**Fig. 10(a).** Consecan offsets and analysis summary



**Fig. 10(b). Received signal level at DSS 14 during Voyager 2 X-band tracking**

PRECEDING PAGE BLANK NOT FILMED



**Fig. 11. Conscan offset signatures from DSS 15 Voyager 2 X-band Conscan snap-on tests:**  
**(a) snap-ons, one axis offset at a time; (b) snap-ons, both axes offset simultaneously**

## Appendix

### Software Programs Used in Antenna Pointing Calibration Work

The software programs used in our antenna pointing calibration work are listed below. Some of the programs existed when the work was started; some were prepared as the needs arose.

At the time of the work, none of the programs had formal status in the operational DSN. Currently, the process is underway for adapting and formally qualifying an appropriate set of the programs for operational support use by the DSN.

The programs are listed in the normal order of their use.

**SCHEDULE.EXE** A. Rius, MDSCC. FORTRAN. Runs on IBM-PC-XT. Produces a radio star tracking schedule with minimum delay between observations.

**PDS.EXE** A. Rius, MDSCC. FORTRAN. Runs on IBM-PC-XT. Accepts SCHEDULE.EXE output file, produces antenna ACS predict file for the radio star track in a format usable at the maintenance port of the ACS.

**MOVE-IT.COM** Woolf Software. MS-DOS Assembler. Off-the-shelf communications program to transfer the PDS.EXE predict file to the iPDS microcomputer for uploading into the ACS.

**ACS-PDS** MDSCC. PLM. iPDS utility. Uploads the radio star predict set into the ACS.

**MDA/NAR** MDSCC. Modcomp II utility to collect noise adding radiometer data during a radio star tracking session.

**AGA** B. Wood, GDSCC. dBaseII. Runs on IBM-PC compatible. Utility to convert SCHEDULE.EXE file into AGA input file.

**AGA** MDSCC. FORTRAN. Modcomp II Antenna Gain Analysis program. Accepts MDA/NAR data disk and tracking schedule from PC AGA utility. Produces antenna pointing offset data versus azimuth/elevation and time.

**PHO** A. Rius, MDSCC. FORTRAN. Runs on IBM-PC-XT. Accepts antenna offset data and produces the multiparameter antenna systematic pointing error model.

**APACRCTB.EXE** R. Riggs, JPL. BASIC. Runs on IBM-PC-XT. Converts multiparameter model into APA compatible systematic error table.

**PCPLOT.EXE** Microplot Systems Co. MS-DOS Assembler. Off-the-shelf IBM-PC utility that emulates VT-100 terminal and uploads systematic error table into the APA.

**CAPTURE.BAS** R. Livermore, CDSCC. BASIC. Generic program that logs antenna angles and offsets vs time by monitoring the ACS to APA communications line. Runs on Sharp PC or IBM PC.

**CON2DAT.BAS** R. Livermore, CDSCC. BASIC. Converts Conscan offset data to PHO compatible input format.

**PLOT.BAS** R. Murray, CDSCC. BASIC. Generic program for the IBM-PC that graphically plots antenna offsets versus antenna azimuth or hour angle.

**CSN-ANAL.EXE** R. Riggs, JPL. BASIC. Computes antenna pointing statistics from data collected by the CAPTURE program. Runs on IBM-PC.

**CONRED2.EXE** T. Kuehn, JPL, Pascal. Runs on IBM-PC. Converts hand-entered antenna offset data into a format usable by the PHO antenna error program.

**OFF-HA.BAS** R. Riggs, JPL. BASIC. IBM-PC utility to list antenna offsets versus hour angle from a multiparameter antenna error model.

**PLOT'MDL.CML** T. Kuiper, JPL. COMAL. Utility for IBM-PC that plots antenna offsets versus hour angle from a multiparameter antenna error model.

**TABLES.EXE** R. Riggs, JPL. BASIC. IBM-PC utility. Tabulates antenna offsets in matrix form vs Az and El in 5-deg increments.

**REF-ACS.EXE** R. Riggs, JPL. BASIC. IBM-PC utility. Computes refraction corrections from environmental conditions.

*DMIT*

## Halley Comet Missions

C. Stelzried

TDA Technology Development

L. Efron and J. Ellis

Navigation Systems Section

*DSN operations in support of the recent Halley Comet closest approach are described in introduction of the subsequent articles.*

In man's continual quest to explore and understand the nature and origin of the solar system, 1985 through 1986 may well be noted as the "Year of the Comet." The first in-situ measurements of a comet occurred on September 11, 1985, with passage of the International Cometary Explorer (ICE) through the tail of Comet Giacobini-Zinner 7800 km downstream of the nucleus. ICE threaded the needle as it scored a bullseye by intersecting the neutral sheet bisecting the tail. Encounter operations support provided by the DSN was described previously in six articles in *TDA Progress Report 42-84*, February 1986. DSN Multimission Navigation Team support of ICE was reported on in *TDA Progress Report 42-86*, August 15, 1986.

This unprecedented encounter was followed six months later by the spectacular rendezvous of a fleet of five spacecraft in the vicinity of Halley's Comet during the four-day long period of March 9 through 13, 1986. The Halley armada included Japan's Sakigake (MS-T5) and Suisei (Planet-A), the Soviet Union's Vega-1 and Vega-2, and ESA's Giotto. Twelve days later, ICE also passed within Halley's sphere of influence.

In addition to the aforementioned six spacecraft, two other interplanetary explorers turned their instruments in the direction of Halley's Comet at various times during the year. The long-lived Pioneer 7 and the Venus orbiting Pioneer 12 contributed to the library of space-based measurements made during the Halley exploration period. Table 1 provides a compilation of the time and distance of closest approach to Halley's Comet for all eight spacecraft. For Pioneer 12, which is in orbit about Venus, this table entry represents the closest approach of Halley and Venus.

It has also been an unprecedeted period in the arena of international cooperation between national space agencies and the worldwide scientific community. The International Halley Watch (IHW) and the Inter-Agency Consultative Group (IACG) provided mechanisms and forums which contributed to the successful encounters with both Halley and G-Z. The IACG initiated the Pathfinder concept using on-board instrument pointing angle data and DSN VLBI data obtained during the earlier arrival of the two Vega spacecraft. This improved the

Halley's Comet ephemeris and hence aided in the final Giotto targeting; of all the spacecraft, Giotto had the closest approach to the nucleus. Articles covering the ICE G-Z encounter science results appeared in *Science*, Vol. 232, April 18, 1986. Halley mission encounter science is reported in brief in the aforementioned issue of *Science* and in greater detail within *Nature*, Vol. 321, No. 6067, May 1986, and *Geophysical Research Letters*, Vol. 13, No. 8, August 1986.

Earth-based observations gathered by the Astrometric Network of the IHW were collected at JPL and distributed to both ESA and Intercosmos. Ephemeris development at JPL for both Comet Giacobini-Zinner and Halley's Comet supported ICE and the two Japanese missions. The DSN participated in the support of all eight missions as briefly summarized in Table 2. The following six articles provide further description of these efforts.

**Table 1. Summary of spacecraft close approaches to Halley's Comet**

Mission	Halley closest approach			
	Date (1986)			
	Distance, km	Day	Month	Hour:min (UTC)
Vega-1	$8.9 \times 10^2$	6	Mar	07:20
Suisei (Planet-A)	$151 \times 10^3$	8	Mar	12:08
Vega-2	$8.0 \times 10^3$	9	Mar	07:20
Sakigake (MS-TS)	$7 \times 10^6$	11	Mar	03:18
Giotto	$5 \times 10^2$	13	Mar	23:03
ICE	$28 \times 10^6$	25	Mar	09:30
Pioneer 7	$12 \times 10^6$	20	Mar	23:36
Pioneer 12 (Venus Orbiter)	$40 \times 10^6$	3	Feb	18:20

**Table 2. Summary of DSN support for Halley's Comet mission set**

Mission	DSN role
Giotto	Telemetry, command, and navigation support for near-Earth departure phase and Halley encounter.
Sakigake (MS-TS) Suisei (Planet-A)	Telemetry, command, and navigation support for first ten days after launch and as requested during cruise. Telemetry support during Halley encounter.
Vega 1 Vega 2	Encounter orbit determination using DSN VLBI techniques.
ICE	Telemetry, command, and navigation support from Dec. 1983 lunar swingby through comet-postencounter interplanetary cruise.
Pioneer 7 Pioneer 12 (Venus Orbiter)	Telemetry, command, and navigation support since launch.

July-September 1986

D24-12  
P.S.

## Giotto Mission Support

C. Stelzried

TDA Technology Development

T. Howe

DSN Control Center Operations Section

*This report summarizes Deep Space Network (DSN) support of the Giotto mission to Comet Halley. It describes the support beginning with the prelaunch testing and continues through the post comet encounter support.*

### I. Introduction

A primary objective of the European Space Agency's (ESA) Giotto mission was to detect the nucleus of Comet Halley using a multicolor camera to produce photographs of the surface from a distance of 1000 km with a 50 meter resolution. A second prime objective was to measure the elemental and isotopic composition of the cometary neutrals, ions, and dust particles, using three specifically designed mass spectrometers. A total of ten experiments were flown on the spacecraft to enable these *in situ* measurements.

The Giotto spacecraft is a spin-stabilized vehicle powered by a solar cell array. Communications with the DSN was via low gain antennas during the early mission phase. The de-spun high gain antenna was used during follow-on mission phases for all communications between the spacecraft and Earth. The spacecraft is able to transmit at both 2.3-GHz and 8.5-GHz frequencies and receive at 2.3 GHz. Only the 8.5-GHz downlink was used for the follow-on mission phases.

With a high flyby velocity at Halley's Comet of 68 km/s, the critical active experiment time was on the order of 4 hours. All data were transmitted in real time, without a backup

recorder capability, at a data rate of 46 kbps. Spacecraft time at closest approach was 00:03:00.4 Universal Time Coordinated (UTC), on 14 March, 1986.

### II. Deep Space Network Support

DSN support of the Giotto mission was divided into four major mission phases: Launch, Cruise, Rehearsals, and Encounter. The support began in June 1985 with prelaunch testing and continued through the comet encounter in March 1986.

#### A. Launch and Cruise

For the cruise phase of the mission the NASA Support Plan (NSP) called for 8 passes per week for the first three weeks following launch using the DSN's 34-meter subnetwork. In addition, two navigation campaigns were to be supported. The DSN was to provide for emergency command should the need arise.

Launch support requirements included prelaunch simulations and the launch phase. Horizon-to-horizon support was

\* C - 4

provided for the first seven days, following launch, by the 26-meter and 34-meter DSN subnetworks. Daily passes of four hours duration were requested from launch plus eight days through the end of July 1985.

**1. Prelaunch Testing and Simulations.** The DSN began support of Giotto on June 7, 1985 with Deep Space Station (DSS) 16's support of a premission test. Prelaunch testing continued through July 1, 1985. All three 26-meter stations (DSS-16, DSS-46, DSS-66), participated in the testing series. A total of 11 tests were conducted with the major portion of tests supported by DSS-46, the initial acquisition station.

**2. Launch Support.** The Giotto spacecraft was launched aboard an Ariane-1 launch vehicle from Kourou, French Guiana at 11:23 UTC, on July 2, 1985. The spacecraft was first placed into a standard geostationary transfer orbit. At third perigee, on July 3, 1985, at 19:24 UTC, it was injected into a heliocentric orbit with the firing of the onboard Transfer Propulsion System (TPS).

DSS-46 began the DSN support of the Giotto mission on July 3, 1985. DSS-46, the 26-meter tracking station located in Canberra, Australia, was selected for launch support due to its relatively wide antenna beamwidth and its high antenna slew rate. This station provided for telemetry reception, tele-command uplink, and radio metric data measurements. Data was routed to the European Space Operations Center (ESOC) at Darmstadt, West Germany via the Jet Propulsion Laboratory in Pasadena, California; the NASA Communications (NASCOM) switching center in Madrid, Spain; and the NASCOM switching center at Goddard Space Flight Center in Greenbelt, Maryland.

Subsequent spacecraft passes were supported by the remainder of the DSN's 26-meter subnetwork consisting of DSS-16 at Goldstone, California and DSS-66 at Madrid, Spain with coverage consisting of horizon-to-horizon passes for the first 7 days following launch. The 26-meter stations continued supporting the Giotto spacecraft throughout the month of July into the first week of August. Their support was supplemented by DSS-42, the 34-meter station located in Canberra, Australia and by DSS-12, the 34-meter station located at Goldstone, California. These passes were on a frequency of one per day with a duration of approximately four hours. Tracking support from the 34-meter antenna (DSS-61) in Madrid, Spain was not requested since the European Space Agency (ESA) stations covered that longitude.

**3. Post Launch Cruise Support.** The DSN support of the launch/early cruise phase concluded on August 6, 1985. No further tracking by the DSN was required until September 15, 1985 when the first Navigation Campaign began. The Naviga-

tion Campaign was supported by the DSN's standard 34-meter subnetwork of stations consisting of DSS-12, DSS-42, and DSS-61. The primary purpose of the Navigation Campaign was to provide ESA with radio metric data for comparison with and validation of the radio metric data collected by the ESA tracking network. The tracking coverage consisted of a series of daily double passes during a two-week interval. A double pass was defined as horizon-to-horizon passes at two stations on different longitudes with at least one of the stations located in the southern hemisphere. The DSN provided the combinations of DSS-12/42, and DSS-61/42 for this support. Data was provided to ESOC in the form of Orbit Data Files (ODF) within 12 hours of the end of each pass. The ODF files were generated from the raw radio metric data (2-way Doppler and ranging) collected during each pass. The data were transmitted via NASCOM high speed data lines to Madrid and then forwarded to ESOC via European communications lines.

## B. Rehearsal Support

The DSN supported the preparation for encounter beginning in October 1985. This preparation consisted of a series of three Encounter Rehearsals. These rehearsals used the Giotto spacecraft as a data source and rehearsed the actual sequences that would be followed during the comet encounter. In addition to the rehearsals, training exercises were conducted by the DSN to insure that the network was fully tested and trained for the encounter. There were also a series of data flow tests conducted each time the wideband satellite circuit, to be used for encounter and rehearsal support, was activated. A final pre-encounter navigation campaign was supported just prior to the March 1986 encounter.

**1. Pre-rehearsal testing.** Pre-rehearsal testing was supported by DSS-42 and by DSS-45, the 34-meter high efficiency antenna in Canberra, Australia. The 34-meter stations were used in place of DSS-43, the 64-meter antenna in Canberra. DSS-43 was scheduled to be the prime supporting DSN station for the Halley encounter but was not available due to Voyager II's use of the station for the impending encounter with Uranus. The primary purpose of this testing was to verify the communications interface between JPL and ESOC using the newly activated satellite circuit. Data flowed to ESOC via NASCOM-provided circuits to the Goddard Space Flight Center (GSFC) and at GSFC the data were patched into a 56-kilobit (kbps) satellite circuit for the final transmission leg to ESOC. The tests utilized the station's Telemetry Simulation Assembly (TSA) to generate the encounter data rate of 46.080 kbps. During the first DSS-42 supported pass on October 2, 1985, data were successfully flowed to ESOC, demonstrating the data interface. Additional tests were scheduled through October 12 using DSS-42 and DSS-45 to further test the circuit interface.

**2. Encounter Rehearsal Number One.** DSS-42 successfully supported the first encounter rehearsal on October 13-14, 1985. Although numerous problems caused delays in acquiring receiver and telemetry lock, the station was able to process and forward telemetry data at rates of 5.7 kbps and 46 kbps. The second part of the first rehearsal was supported by DSS-42 on October 15-16, 1985. This test was less than successful in that anomalies experienced during the first test were still evident. The station had difficulties acquiring the main antenna beam and the main carrier. Predictions for Doppler were offset by several thousand hertz. A replay of the data recorded during the test was scheduled for October 17.

**3. Post Rehearsal Data Flow Tests.** Since the results of the first rehearsal test were not as expected, a series of data flow tests was scheduled between October 17 and 23 to troubleshoot the system. The first test was conducted on October 17 and used DSS-42 as a data source. Testing continued on nearly a daily basis with improving results on each test. The post rehearsals tests concluded with a spacecraft pass on October 23. Equipment problems at DSS-45 delayed the establishment of good data flow to ESOC. However during the final test period, data were 99.7 percent error free.

**4. Circuit Re-activation Data Flow Testing.** In November 1985, the DSN was notified by the Giotto project that the second navigation campaign, originally scheduled for December 1985, would not be required. The time previously scheduled for the campaign was used for data flow testing following the re-activation of the wideband satellite circuit between GSFC and ESOC.

The primary purpose of this series of tests was to:

- (1) Validate and re-certify the communications interface between the station and ESOC.
- (2) Test various configurations including ground communications error correction, and establish the encounter support configuration with the lowest throughput error rate.

The first tests were conducted on December 11, 12, and 13. The tests used a station digital recording with known good data quality, recorded during the October tests, as a data source. These tests determined that error correction was not a viable function for encounter support due to the expected high data rate of 46 kbps. It was also established that data from the two overseas complexes in Spain and Australia would be routed to GSFC and then directly to Europe via the satellite circuit. This configuration would reduce the transmission path by approximately 6000 miles and reduce the number of transmission errors. The overall data flow configuration is shown in Fig. 1.

Testing continued on December 16, 17, 18, 19, 22, and 29. These tests used a combination of recorded data playbacks and spacecraft passes to continue verification of the data delivery quality to ESOC in preparation for the second encounter rehearsal to take place in January 1986. These tests were all considered successful with system refinements being accomplished with each test.

**5. Encounter Rehearsal Number Two Support.** The second in a series of three encounter rehearsals was conducted on January 3, 6, 9, 12 and 15, 1986. DSS-42 provided DSN support for this rehearsal. DSS-43 was supporting the Voyager II Uranus encounter and was not available for Giotto support. The standard encounter support configuration was utilized with the spacecraft providing the data source. All DSN support was in parallel with the 64-meter antenna support provided by the radio astronomy facility at Parkes, Australia.

This series of passes resulted in the refinement of the procedures and techniques to be used for the encounter. Special attention was directed towards the switching and monitoring of data at the GSFC NASCOM switching center, the refinement of station predictions, monitoring of data quality as seen at JPL, a procedure for determination of downlink frequency while three-way with the Canarvon, Australia tracking station, and the localization of potential problem areas in the encounter configuration.

In January the DSN was notified of additional Giotto support requirements. As a result of the findings during the International Cometary Explorer (ICE) flyby of Comet Giacobini-Zinner it was determined that the data-taking period for a comet encounter should be greatly expanded. The Giotto project requested three days of continuous 64-meter antenna coverage beginning on March 12, 1986. The 64-meter stations in California, (DSS-14), Australia (DSS-43), and Spain (DSS-63) would provide this support. In addition to the expanded encounter support, the DSN was also requested to support a final navigation campaign during the first week in March 1986. The purpose was to refine the targeting for the comet flyby. The bulk of the campaign support was to be provided by DSS-12 at Goldstone, California.

**6. Giotto Spacecraft Emergency.** On January 24, 1986 the DSN was notified of a Giotto spacecraft emergency following the loss of downlink lock during a pass being supported by the German tracking station. Since the project had no visibility into the health of the spacecraft, they were greatly concerned about the possibility of the spacecraft overheating due to a possible misorientation. The project requested DSN support and was given DSS-12. This station was able to lock up to the spacecraft carrier, however the telemetry remained marginal due to the low downlink signal level. DSS-12 was able to pro-

vide telemetry with an error rate of about 50 percent. The 64-meter station at Goldstone was then requested. DSS-14 was supporting the Voyager II Uranus encounter at the time. However, following negotiations between the Giotto and Voyager project managers the station was turned over to Giotto. Even though this was the first support of Giotto by DSS-14, they were able to acquire the spacecraft and flow data to ESOC in about 37 minutes. Following the analysis of engineering telemetry it was determined that the spacecraft's high gain antenna was mis-pointed. A corrective command load was generated and was successfully transmitted to the spacecraft by Canarvon, Australia as soon as the spacecraft came into their view.

**7. Support of Encounter Rehearsal Three.** The third in the series of encounter rehearsals was supported on 5, 7, 9, 11 and 13 February, 1986. Unlike the previous rehearsals, this series was to be supported by DSS-43 as well as DSS-42. DSS-42 supported the passes on the 5th, 7th, and 13th while DSS-43 supported on the 9th and 11th. As in the other rehearsals, the goal of the test was to simulate as closely as possible the sequence of events to be followed for the actual encounter. The spacecraft was the data source.

The first four passes of this rehearsal were supported by the DSN nearly flawlessly. During the final rehearsal pass on February 13, DSS-42 data was of poor quality at ESOC following an apparent spacecraft modulation index change. ESOC verified that this change had not been observed at Canarvon. A troubleshooting effort at both the station and ESOC could not correct the problem and the last three hours of data were lost to the project. Extensive troubleshooting and investigation following the pass could not pinpoint the source of this outage. It was learned, however, during a playback of the data recorded during the period in question, that the data recorded on the digital tape at DSS-42 was of good quality. A communications buffer was subsequently changed as the most likely trouble source. This problem did not recur.

**8. Uplink Handover Demonstration.** In order to demonstrate the DSN's ability to transfer uplink control of the spacecraft from an ESA to a DSN station, a test was conducted on January 22, 1986. The test showed that the uplink could be transferred between Canarvon and DSS-43 without difficulty. The DSS-43 uplink for command would be required during the encounter only in emergency support cases.

**9. SPC 10 and SPC 60 Encounter Training.** With the expanded DSN 64-meter subnetwork coverage requirement for the encounter, the DSN found it necessary to conduct additional training exercises with the Signal Processing Centers at Goldstone (SPC 10) and Canberra (SPC 40). The Giotto spacecraft was used as a data source and with the project's coopera-

tion the data rate was commanded to 46 kbps so that the encounter rate could be processed. Six tests were conducted with each complex. Additional training was accomplished as a result of the complex's support of the final navigation campaign in March.

**10. Final Encounter Rehearsal.** The last two tests in the series of encounter rehearsals were conducted on February 28 and March 10, 1986. DSS-43 supported these tests. In addition to processing spacecraft telemetry at the encounter rate of 46 kbps, uplink handovers were also demonstrated between DSS-43 and Canarvon.

### C. Encounter

**1. Final Navigation Campaign.** The last navigation campaign prior to the encounter was conducted from March 2 through March 12, 1986 and consisted of a total of 15 passes using DSS-12, DSS-42, and DSS-63 to obtain two-way coherent Doppler and ranging. The only problem encountered during this campaign was the failure of DSS-12 to obtain uplink lock during one pass. It was speculated that the high uplink power level being used by station 12 may have caused the failure to lock. A reduction in power level was requested during subsequent passes and the problem did not recur.

**2. Encounter Support.** Encounter Support by the Deep Space Network officially began on March 11, 1986 at 1610 (UTC). Coverage was for three continuous days by the 64-meter subnetwork. In addition to supporting the Giotto spacecraft, DSS-42 supported the Japanese MS-T5 (Sakigake) Halley Comet mission spacecraft on encounter day to collect radio metric data for calibration of the earth's atmosphere. After a nearly uneventful eight-month voyage through space, the Giotto spacecraft was about to begin the critical data-taking period. During the four hours leading up to the encounter, which took place at 00:03 UTC, spacecraft time, on March 14, 1986, the ten on-board experiments began to receive data. The closure speed between the spacecraft and the comet at that time was on the order of 68 km/sec. At 23:01 UTC the project reported the first dust impacts at a distance of about 287,000 km.

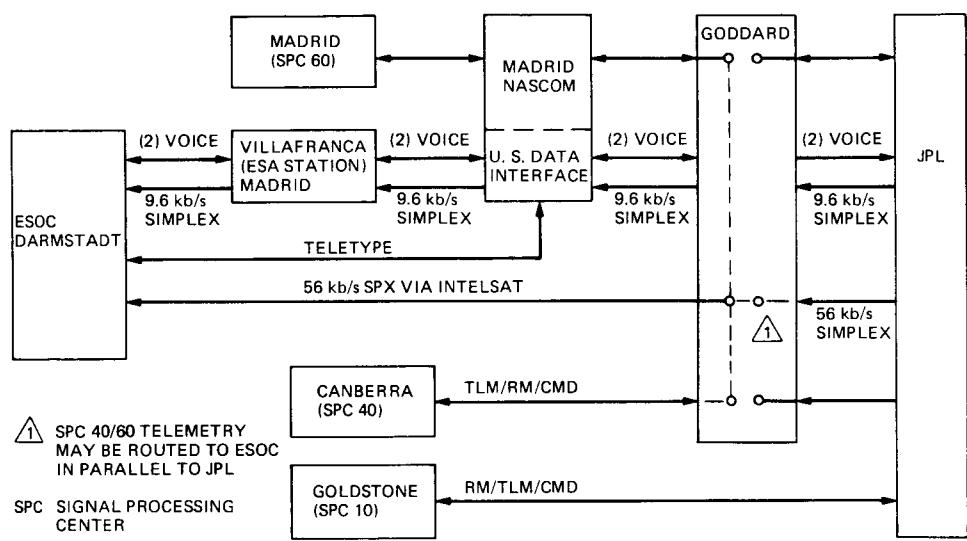
The DSN and spacecraft functioned perfectly up until two seconds before the time of closest approach when the spacecraft signal was lost due to dust particles impacting the spacecraft and causing nutation of the spin axis. The nutation caused the spacecraft's high gain antenna to offpoint and the downlink was lost. The spacecraft's nutation dampers slowly compensated for this effect. Receiver lock was restored after about 20 seconds; however, it took nearly 34 minutes for the real time telemetry to be fully restored. Analysis is still underway to fully understand this phase of the mission. Attempts

are being made to recover data from this marginal period using the station's digital recordings and special recovery techniques. The project reported that the overall encounter results were spectacular. Pictures had been obtained to about 1500 kilometers from the nucleus before the signal was lost and had passed within 590 km of the nucleus and survived. Although

post encounter assessments showed the spacecraft has received damage as a result of dust particle impacts, ESA made the decision to perform a maneuver which will allow Giotto to pass by the Earth in July 1990 for a possible retargeting to another comet. The spacecraft is now in a hibernation mode; DSN support of Giotto concluded on March 15, 1986.

## References

1. Reinhard, R., "The Giotto Encounter With Comet Halley," *Nature*, Vol. 321, pp. 313-318, 15 May 1986.
2. Wilkins, D. E. B., Parkes, A., and Nye, H., "The Giotto Encounter and Post-Encounter Operations," *ESA Bulletin*, Number 46, pp. 66-70, May 1986.



**Fig. 1. Giotto DSN/ESOC data flow configuration**

# Giotto Navigation Support

N. A. Mottinger and R. I. Premkumar  
Navigation Systems Section

*Cooperative efforts between NASA and the European Space Agency (ESA) in supporting the flight of Giotto to Halley's Comet included prelaunch checks of ESA navigation software and delivery of validated DSN radio metric tracking data during the mission. Effects of drag from passing through the coma are seen in data received pre and post encounter. The post encounter Giotto trajectory provides a solar occultation in January 1988, prior to returning to the earth in 1990 for possible retargeting to yet another comet.*

## I. Introduction

NASA, through the DSN, played a vital role in the navigation of the European Space Agency's (ESA) Giotto spacecraft to Halley's Comet by receiving radio metric tracking data from the spacecraft and transmitting it to the European Space Operation Center (ESOC) in Darmstadt, West Germany. Insuring that ESOC was prepared to use the data for its first deep space mission involved several years of detailed checks of their navigation program. At the same time, communication systems needed to transfer information enabling the DSN to acquire the spacecraft and ultimately return validated radio metric data in a timely fashion to ESOC were defined and checked out.

This report summarizes the checkout procedures used in preparing ESOC software for navigating Giotto with DSN tracking data and describes the flow of information between JPL and ESOC during the mission. Although JPL was not required to provide estimates of the spacecraft encounter, a brief comparison of the ESOC and JPL derived encounter conditions is included. Also shown is evidence of the drag experienced by the spacecraft while passing through the comet's coma.

The mission has not ended. Giotto survived the flyby, not unscathed, but intact enough for retargeting to the earth and perhaps eventually to another comet. In the interim, opportunities to probe the solar corona, including a solar occultation are present. Plans for using VLBI techniques to improve orbit determination and hence the occultation science return, may be tested using Giotto in early 1987.

## II. Pre-Flight Navigation Activities

Beginning more than two years before the Giotto launch a series of navigation software workshops were held between JPL and ESOC to define and run test cases to verify that the ESOC orbit determination program could successfully process Giotto radio metric data. The tests concentrated on basic orbit determination functions:

- (1) Integration of the spacecraft trajectory and variational equations.
- (2) Light time solution, time transformations and polar motion.
- (3) Computation of observables and partial derivatives.
- (4) Differential correction, covariance matrix and mapping.

These functions were tested using the Voyager 1 trajectory and DSN radio metric data acquired from it when the geometry was similar to the forthcoming Giotto encounter. This tracking data was initially sent to ESOC by magnetic tape and later transmitted over communication lines as tests of the system to be used for sending DSN radio metric data during the mission.

An important part of the software tests involved the choice of a planetary ephemeris. It is not only the source of position and velocity of bodies in the solar system, but also of nutation and precession of the earth and a host of astrodynamical constants such as body masses, the length of the astronomical unit and the speed of light. It defines the coordinate system for the dynamics of the spacecraft flight and dictates the values of station locations required to properly process radio metric observables. The one chosen for Giotto operations and hence these tests was JPL Development Ephemeris (DE) 118, which uses the Earth Mean Equator and Equinox of 1950 reference system. It would be the source of data for all solar system bodies other than Comet Halley, which was of no immediate consequence in these tests.

Serving as a standard of comparison would be the JPL Orbit Determination Program (ODP) (Ref. 1) first used in support of Mariner VI and VII and all space missions tracked by JPL since. The test cases were run on the JPL ODP and compared with the same case run with the ESOC ODP. All the JPL cases were run on a UNIVAC 1100 computer with a double precision word length of 18 decimal digits. ESOC used a SIEMENS computer with 16 decimal digits double precision. For Giotto operations JPL switched to a VAX 780, which has the same word length as the SIEMENS. Tests between the two JPL computers showed no navigation degradation for a Giotto type trajectory due to the shorter word length.

The testing began by matching the integration of the spacecraft trajectory between the JPL and ESOC programs. The reference trajectory was based upon a 2.5 month long Voyager 1 trajectory modified to include large spacecraft maneuvers and non-gravitational accelerations to enhance the detection of any possible differences between the two programs. Good agreement between ESOC and JPL was noted at 1 meter in position and 1.E-5 m/s in velocity at the end of the integration.

Two Voyager 1 Doppler points were selected for use in a detailed computation check. Quantities carefully compared were time transformations, polar motion, light time solution, EME50 station location, antenna corrections, and troposphere modeling. The final agreement obtained for the computed observables was 0.0001 Hz S-band or approximately 0.007 mm/s,

which indicated that the ESOC ODP could process DSN radio metric data adequately to support Giotto navigation.

Checking the partial derivatives of the observable with respect to spacecraft state, station locations, and polar motion parameters could not easily be done due to differences in the formulation of the filters of the JPL and ESOC programs. The ESOC program uses a current state filter while JPL uses an epoch state filter. Although it would have been possible to map the JPL partials to coincide with the formulation of the ESOC program, this was not done when good agreement between the programs was obtained in the solution for the spacecraft state programs (noted below).

A more comprehensive test of the ESOC program involved the estimation of spacecraft position and velocity using an eight day span of Voyager data. Differences in the estimates obtained by the two programs were 22 km in position and 20 mm/s in velocity. In view of Giotto navigation accuracy requirements of approximately 100 km (1-sigma), exclusive of errors in the comet ephemeris, these differences were considered acceptable for successful navigation. A portion of the difference might be attributed to data processing techniques. For example, polynomial representation of the data smooths it in a preprocessing step before use in the ESOC ODP. Differences between the epoch state filter used in the JPL ODP and the current state estimator employed in the ESOC program might also be a contributing factor although attempts were made to match the filters as closely as possible.

Mapping tests conducted using the above estimation case, involved only a translation in time without changing coordinate systems. These showed the same type of agreement as noted above. Complete test results are reported in an ESOC document, "Giotto Quality Control Document, Single Tests, Results of Orbit Determination Test Runs on Voyager Data", Document GIO-QCD-3, Issue no. 2, F. Hechler and H. Muller, European Space Agency Operations Center, 15 March 1985.

Another orbit determination test involved a 90-day arc of radio metric data. The data was based upon the geometry of the Giotto-like Voyager 1 trajectory, but for extended software checking purposes used precisely known solar pressure, instantaneous, and finite maneuvers. Also present during the last month was a constant acceleration (gas leak). The data observation model also included troposphere effects and random data noise representative of that expected when the DSN would track Giotto. Since the trajectory was known exactly, it was possible to determine when the correct state had been recovered by the ESOC ODP in runs using intentionally mismodeled a priori spacecraft states and/or non-gravitational force models. The ability to recover the correct spacecraft

state was comparable to or better than the agreement obtained in the previous Voyager check.

The results of these tests indicated that ESOC should be able to process DSN radio metric data acquired from Giotto. This was indeed confirmed during the mission when differences of 15 km and 17 mm/s in position and velocity, respectively, were noted between JPL and ESOC reductions of DSN data.

### III. Flight Operations Activities

#### A. Navigation Campaigns

The DSN was available under contract to ESA for supplying radio metric data at pre-arranged times during the Giotto mission. Two week tracking periods comprised primarily of two successive DSN passes per day intermingled with tracking collected at the ESA Deep Space Tracking System (DSTS) were called campaigns. These served to enhance Giotto navigation, primarily during the approach and encounter phase by supplying Doppler and range from northern and southern hemisphere sites. There were two stations in the DSTS, a 15-meter antenna at Carnarvon in western Australia and a 30-meter one at Weilheim in the Federal Republic of Germany.

The first navigation campaign occurred from 15 to 29 September 1985, a period during which there were no spacecraft maneuvers of any kind. DSN radio metric tracking was obtained from the 34-meter stations, DSS 12, 42, and 61, and was transmitted to ESOC once a week using the NASCOM network. Table 1 summarizes some information about this data after processing with the JPL ODP and Fig. 1 shows Doppler residuals obtained from both JPL and ESOC processing and likewise Fig. 2 shows range residuals. The Doppler data were compressed to a 600-second count time. The residuals obtained by the two programs are very similar. Some of the differences may be due to variations in data processing techniques. For example, ESOC weights the range at 10 meters whereas JPL weights it at 1000 meters to reduce sensitivity to unmodeled errors, namely range biases and/or station location errors. ESOC may have also processed all the range data collected, while the JPL analysis restricted the range to one point every one-half hour. Other factors possibly contributing to the small differences include the pre-processing of the tracking data required to change it to a polynomial representation for use in the ESOC ODP.

The marked similarity in results served as another verification of ESOC's ability to process DSN radio metric data, as expected after all the software testing, and established confidence for successful data processing during the critical encounter campaign when DSN data would again be received.

Other material presented by ESOC at a navigation workshop following the first campaign included residuals of the combined DSN and DSTS tracking data. To successfully process data from both tracking networks it was necessary to determine the location of the DSTS stations in the coordinate system defined by the planetary ephemeris. Planetary Ephemeris DE 118 had previously been used to determine the locations of the DSN stations thereby enabling ESOC to estimate the DSTS locations while holding the DSN locations fixed. These locations were then used for the rest of the mission, which would be especially significant during the encounter campaign when radio metric data from the two networks would again be combined for determining the orbit and designing the final trajectory correction maneuver for a successful encounter.

The second navigation campaign, supporting the encounter, occurred from 1 to 12 March 1986 in which DSTS radio metric data were again augmented with daily DSN radio metric tracking obtained from the 34-meter stations, DSS 12 and 42, and the 64-meter stations, DSS 14 and 63. At the end of a day's tracking, data were validated with the JPL ODP and transmitted via NASCOM to ESOC. Minimal comparison of ESOC and JPL data analysis occurred due to the high activity associated with this phase of the mission. A change in the spacecraft transponder configuration implemented much earlier in the cruise phase caused the received 2-GHz (S-band) signal to be retransmitted at 8.5 GHz (X band). Table 2 shows a summary of the JPL processing of the DSN data taken during this campaign. The noise of the Doppler appears higher than for the first campaign due to two passes of 60-second count time data that could not be compressed. Figures 3 and 4 show plots of the two-way Doppler and range residuals obtained with the JPL ODP.

There was not a requirement for the exchange of navigation results between JPL and ESOC during either of the campaigns. The exchange following the first campaign occurred several months afterwards and served only as a final check of ESOC's ability to successfully process DSN and DSTS radio metric data. The DSN data taken during the encounter campaign aided ESOC in estimating the trajectory of the spacecraft, now that daily attitude maneuvers were occurring, in preparation for the design of the final trajectory correction maneuver. The last DSN pass of this campaign followed that maneuver and helped provide verification of the achieved trajectory change.

#### B. Navigation Campaign Data Processing

Several interfaces were required between JPL and ESOC in order for the DSN to successfully track Giotto and transmit validated tracking data to ESOC. One of the unique features of the system devised was the capability to directly access data stored in the operations computer at ESOC from JPL.

DSN antenna pointing predicts, required to acquire the spacecraft signal, were generated from a Giotto trajectory integrated by the JPL ODP using initial conditions and solar pressure model obtained from a file in the ESOC computer. Although the conditions in this file were updated weekly, it was not necessary to actually update the predicts very often. In spite of daily attitude maneuvers and the final pre-encounter trajectory correction maneuver, analysis with these current states indicated that predicts generated from a December 11, 1985 state and solar pressure model were adequate to support the project throughout the remainder of the mission, including the encounter phase.

Validation of the DSN radio metric tracking data was performed using a Giotto trajectory generated from a current state and maneuver information obtained from ESOC computer files. The tracking data were prepared by compressing the Doppler to an ESOC specified count time and verifying the use of the correct station and spacecraft hardware delays for the range data. The JPL ODP was then used to validate the data, noting any blunder points to remove before transmission to ESOC.

The NASA Communications Network, NASCOM, was used as part of a communications network to transmit the validated data from JPL to ESOC where it was routed to the operations computer for use in navigation.

### C. Encounter Estimate

An estimate of the spacecraft arrival time and position at Halley was derived at JPL from DSN radio metric data collected during the second campaign. Although there was no requirement to deliver these estimates to ESOC, this was done for our own information and then compared with the value obtained by ESOC. Due to the many facets associated with the daily attitude maneuvers and possibly other information known only at ESOC about the attitude behaviour of the spacecraft, one might not expect close agreement between the two solutions. Cause for disagreement in the solutions could also easily come from differences in the comet ephemerides used.

The JPL solutions were obtained with a simple least-squares batch filter, while modeling the maneuvers using data from the ESOC maneuver file. After examining the effects of estimating various sets of the daily attitude maneuvers and the trajectory correction maneuver (TCM), which occurred on 12 March around 01:30 UT, in combination with different a priori uncertainties, the choice was made to estimate 10 of the 12 daily maneuvers with an a priori uncertainty of 10 cm/sec and the TCM with an a priori uncertainty of 50 cm/sec. Data statistics for the residuals obtained from this solution are those shown previously in Table 2. JPL Planetary ephemeris DE 118 and

the International Halley Watch comet ephemeris HL47 were used. This Halley ephemeris is derived from earth-based observations ending 24 March 1986 and includes a center of light center of mass offset. The gravitational effect of the comet on the trajectory of the spacecraft was ignored. The analysis used the consider option to augment the covariance of the estimated spacecraft position for the effects of possible errors in the tracking station locations of 2 meters in distance from the spin axis, 3.0 E-5 degrees (approximately 3 meters) in longitude, and 20 meters in distance from the equator plane.

A predicted comet miss distance of 610 km with an uncertainty of 104 km was obtained. Augmenting this uncertainty for the assumed station location errors resulted in an uncertainty of 138 km. The predicted time of closest approach is 14 March 1986, 0<sup>h</sup> 2<sup>m</sup> 58.5<sup>s</sup> UTC with an uncertainty of 0.8 s which grows to 1.1 s with the consider parameters.

Analysis by ESOC (Ref. 2) using both pre and post encounter DSN and DSTS radio metric data and their comet ephemeris derived from Earth-based observations augmented with the Vega-1 and Vega-2 Halley observations also indicates that the actual miss distance was 610 km with an uncertainty of 40 km with a time of closest approach of 0<sup>h</sup> 3<sup>m</sup> 0.4<sup>s</sup> UTC.

### D. Passing Through the Halley Coma

A drag effect attributed to Giotto passing through the comet's coma can be observed in pre and post encounter two-way DSN radio metric data. Figures 5 and 6 show changes in Doppler and range residuals obtained when an ESOC provided post TCM Giotto state was integrated forward and used to form residuals. The first post encounter data was obtained about 12 hours after closest approach and shows an offset of 9.4 Hz, X-band or 171 mm/sec in the Doppler, Fig. 5. This is independently confirmed from the slope in the post encounter DSS 14 range data, Fig. 6, which yields similar velocity change of 168 mm/s. (There was no pre-encounter range taken following the TCM.) A value for the total reduction in the spacecraft velocity can then be computed knowing that the earth direction is 44.2 deg from the velocity vector. The value derived for the total velocity, 238.5 mm/sec, agrees to within 4 mm/s with estimates made by T. A. Morley at ESOC from DSTS two-way tracking ("Braking Effect of Dust Impacts on GIOTTO at Encounter," T. A. Morley, European Space Operations Centre, 18 March 1986). Simple assumptions of inelastic impacts along the velocity vector and the conservation of momentum infer the total mass of the impact dust to be 2.0 grams. This is obtained assuming the velocity of the impacting dust to be 68.377 km/s and using the ESOC value (Morley) for the spacecraft mass of 573.886 kg. More thorough analysis by Morley using longer spans of DSN and DSTS pre and post encounter two-way radio metric data gave an estimate of 1.9 g.

Analysis of this same data by the Giotto Radio Science Team reported in Ref. 3 indicates that large uncertainties in the momentum multiplication factor arising from enhanced momentum transfer due to inelastic high-energy particle impacts reduces the total mass of the impacting dust to the 0.1–1 g. range.

The two-way radio metric data analyzed above was collected several hours either side of encounter and therefore cannot be used to probe the nature of the coma itself. One-way Doppler received throughout the encounter period, could be used for studies of the coma, but requires substantial analysis to extract meaningful information. Figure 7 shows three minutes of this data recorded at DSS 43. Some signatures in this data correspond with known particle impacts, but cycle slips are most certainly present and lock was lost during this interval. These all probably invalidate the large offset of approximately 17 Hz observed pre and post encounter. This offset is about a factor of four greater than the effect observed in the two-way Doppler discussed above which indicates the need for careful interpretation of this data. It appears that if any information about the coma is to be gleened from the one-way data, it will require analysis of the open loop recordings, a task which is currently underway. One-way observations at Carnarvon and Parkes, Australia reveal similar signatures. There do not appear to be any plans to continue the analysis of these data.

#### **IV. Post Encounter Trajectory and Possible Future Activities**

Following the encounter a series of spacecraft maneuvers were performed which placed Giotto in a trajectory which would fly by the earth at about 20,000 km in July 1990 for subsequent retargeting to another comet. Following these

maneuvers, daily communication from the DSTS to Giotto ceased, DSN communication having previously ended following the DSS 14 pass on 14 March. Giotto is in a hibernation state with only occasional communications planned (Ref. 4). The resulting trajectory contains an extended period of some 150 days during which the angular separation of Giotto and the sun will be less than 10 solar radii climaxing with a 5 day solar occultation in January 1988. Figure 8 shows Giotto relative to the sun during this 150-day period in a coordinate frame in which the trace of the Giotto trajectory as seen from the earth is plotted in a plane located at the sun and perpendicular to the fixed earth-sun line. The axes of the plot show the angular separation in right ascension and declination of the spacecraft from the sun. A perspective of the trajectory throughout the entire hibernation period can be seen in Fig. 9 in the same type of a plot. The horizontal axis, which serves as an approximate sun-earth-Giotto angle, indicates that Giotto is always within 70 deg of the earth-sun direction.

Two additional plots of general interest are the geocentric declination, Fig. 10 and right ascension, Fig. 11, covering the time span November 1986 to June 1988. Note that the solar occultation occurs near -23 deg declination while the June solar graze occurs at about +23 declination.

In anticipation of the interest surrounding the solar occultation and the need for accurate orbit determination in the presence of fairly frequent maneuvers, a VLBI experiment using Giotto is being studied for early 1987. Preliminary analysis indicates that this data combined with conventional radio metric observables can be very effective in determining the orbit in this environment. Results of this proposed experiment should be of interest to the Ulysses project which uses the same radio transponder and will be probing the solar environment also.

## References

1. Moyer, T. D., *Mathematical Formulation of the Double-Precision Orbit Determination Program (DPODP)*, Technical Report 32-1527, Jet Propulsion Laboratory, Pasadena, Calif., May 15, 1971.
2. Munch, R. E., Sagdeev, R. Z., and Jordan, J. F., "Pathfinder: Accuracy Improvement of Comet Halley Trajectory for Giotto Navigation," *Nature*, Vol. 321, No. 6067, pp. 318-320, 15 May 1986.
3. Edenhofer, P., Bird, M. K., Brenkle, J. P., et al., "First Results From the Giotto Radio-Science Experiment," *Nature*, Vol. 321, No. 6067, pp. 15-21, May 1986.
4. Morley, T. A., and Hechler, F., "GIOTTO, Orbit Prediction During Giotto Hibernation," European Space Operations Centre, *GIOTTO Study Notes*, No. 60, April 1986.

**Table 1. DSN data summary for Campaign #1**

A. Amount of Tracking Data Processed					
Data Type	Number of Points			From	To
	Received	Used	% Used		
Doppler (F2)	532	532	100	15 Sept 20:59	29 Sept 11:51
Range (PLOP)	511	139	27	15 Sept 20:53	29 Sept 11:44 (Actual amount used reduced to facilitate processing.)

B. Total Amount of Tracking Data Received			
Station ID	Data Type	Band <sup>a</sup>	Points
DSS 12	F2	S	138
DSS 42	F2	S	303
DSS 61	F2	S	91
DSS 12	PLOP	S	139
DSS 42	PLOP	S	298
DSS 61	PLOP	S	74

C. Data Statistics and Weights			
Data Type	Residuals		Data Weight
	Bias	Sigma	
Doppler (F2)	0.00001 mm/s	0.074 mm/s	1 mm/s (60 s count)
Range (Plop)	0.50 m	25 m	1000 m

**Table 2. DSN data summary for Campaign #2**

A. Tracking Data Received					
Data Type	Number of Points			From	To
	Received	Used	% Used		
Doppler (F2)	1008	955	95	1 March 12:46	12 March 18:52
Range (PLOP)	133	131	98	1 March 13:43	11 March 12:08

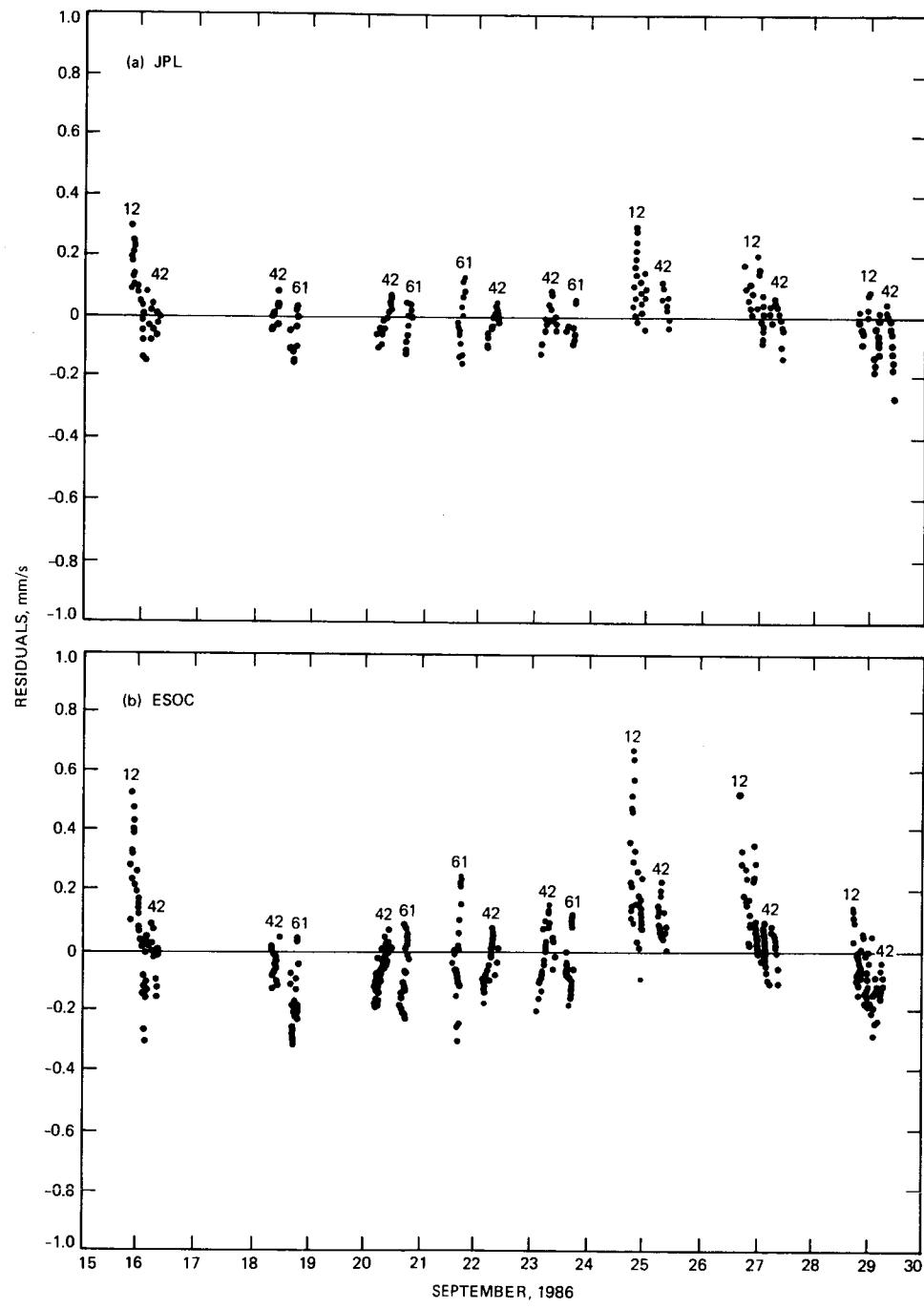
B. Total Amount of Tracking Data Received			
Station	Data Type	Band <sup>a</sup>	# Points
DSS 12	F2	S/X	792
DSS 14	F2	S/X	36
DSS 42	F2	S/X	96
DSS 63	F2	S/X	31
DSS 12	PLOP	S/X	86
DSS 42	PLOP	S/X	36
DSS 63	PLOP	S/X	9

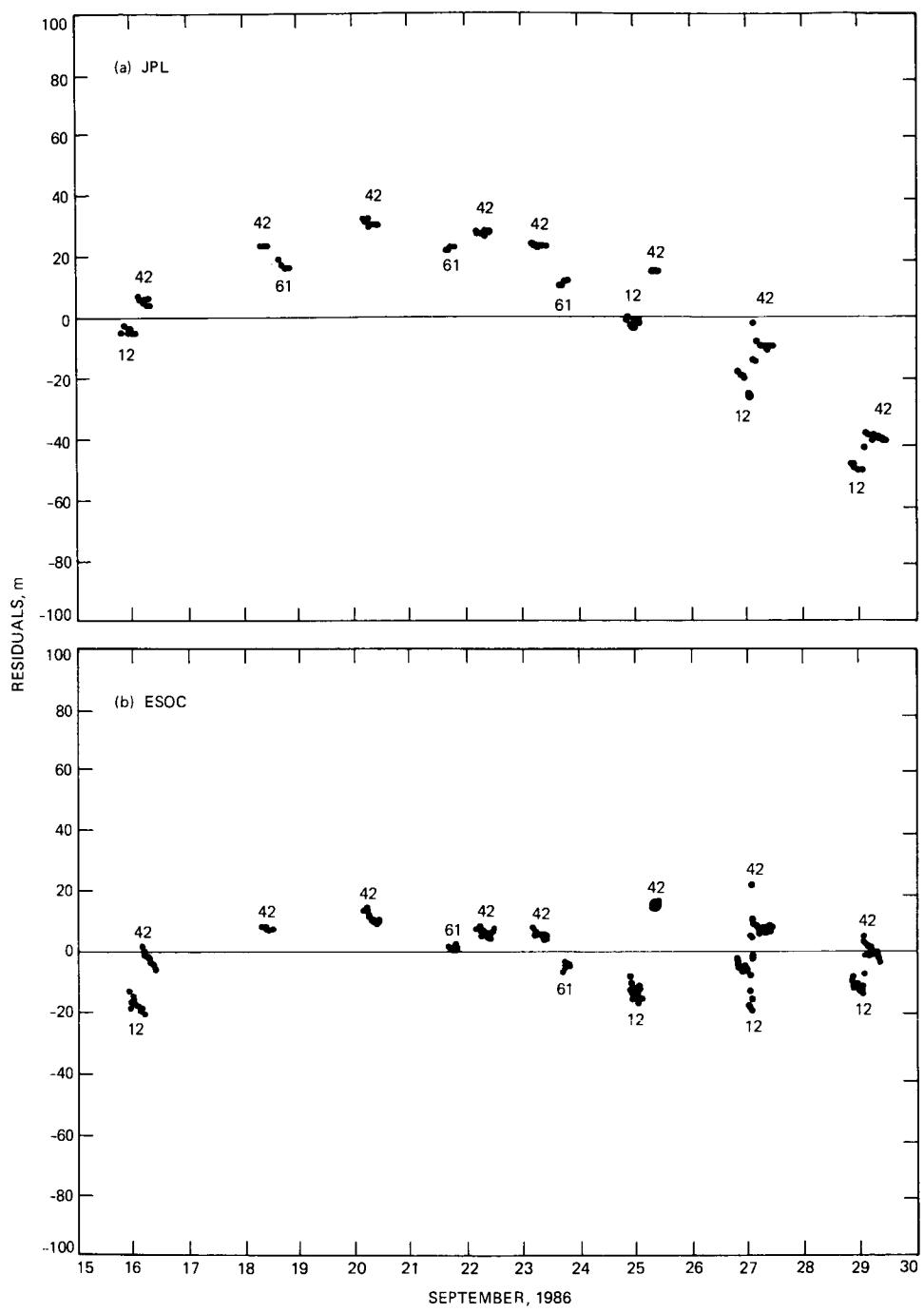
<sup>a</sup> S = 2 GHz; X = 8.5 GHz
-------------------------------------

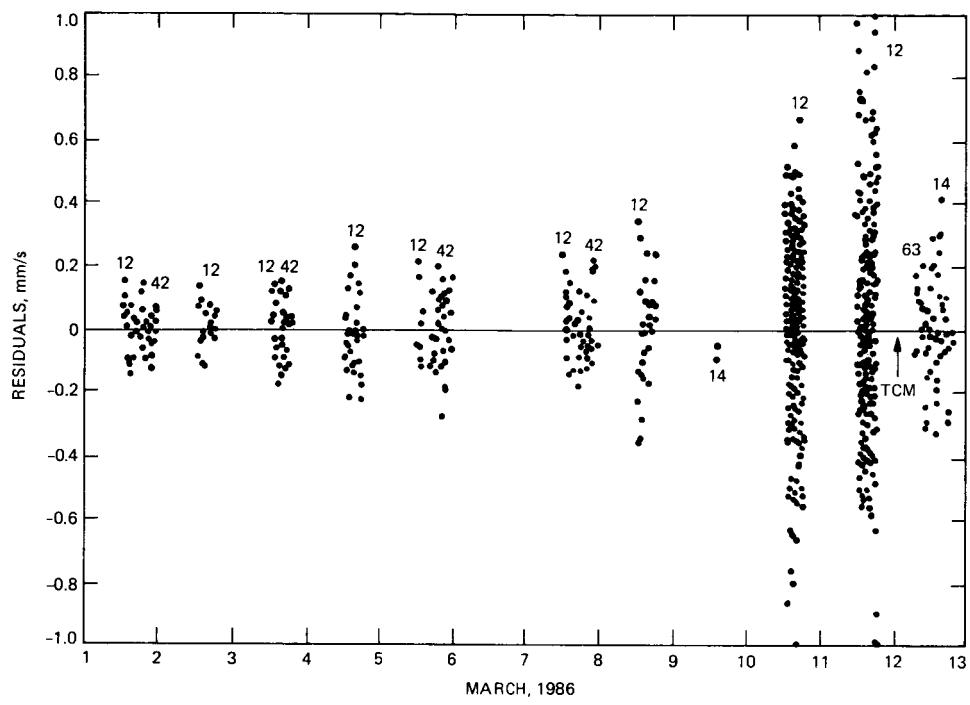
C. Data Statistics and Weights			
Data Type	Residuals		Data Weight
	Bias	Sigma	
Doppler (F2)	0.00089	0.30 mm/s	1 mm/s (.055 Hz)
Range (PLOP)	-0.095	29 m	1000 m



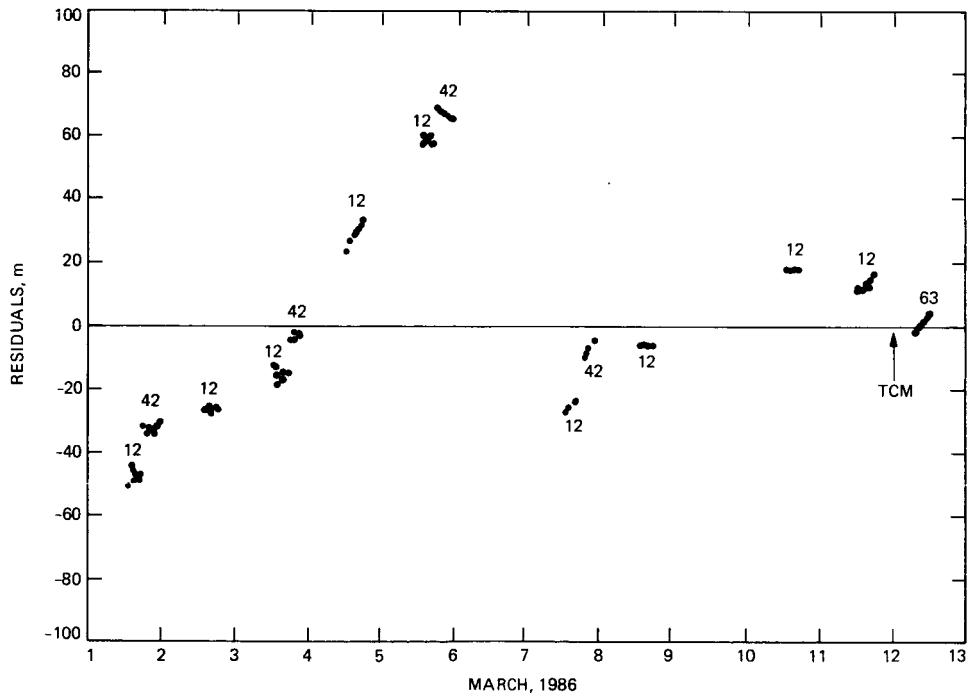
**Fig. 1. JPL and ESOC DSN Doppler residuals from Campaign #1**



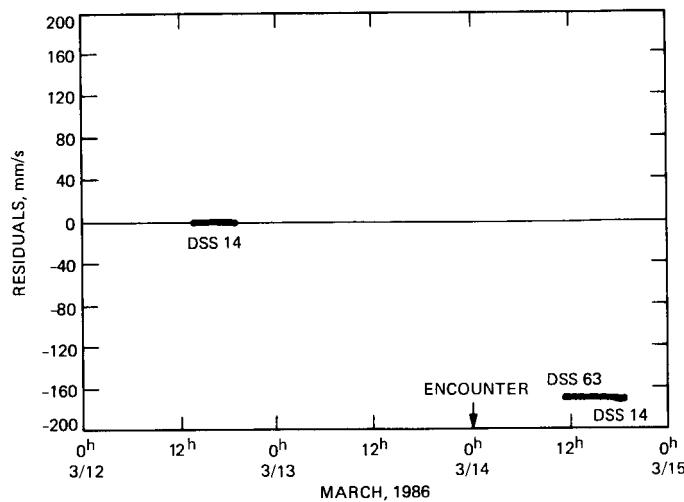
**Fig. 2. JPL and ESOC DSN range residuals from Campaign #1**



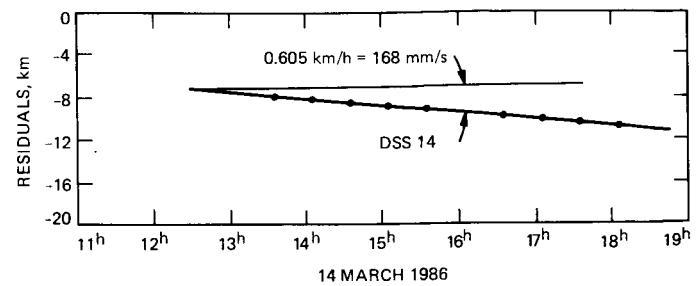
**Fig. 3. JPL DSN Doppler residuals from Campaign #2**



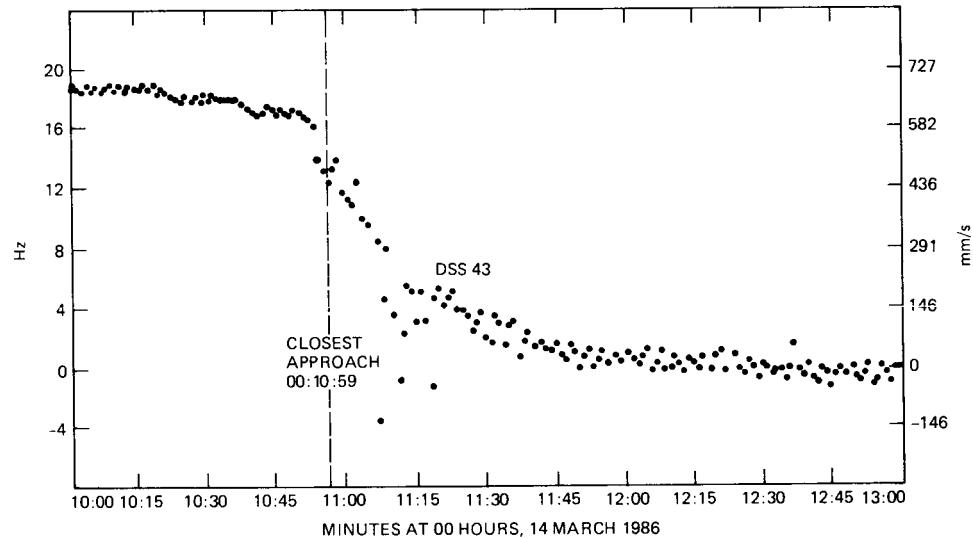
**Fig. 4. JPL DSN range residuals from Campaign #2**



**Fig. 5. Post encounter velocity change observed in Doppler residuals**

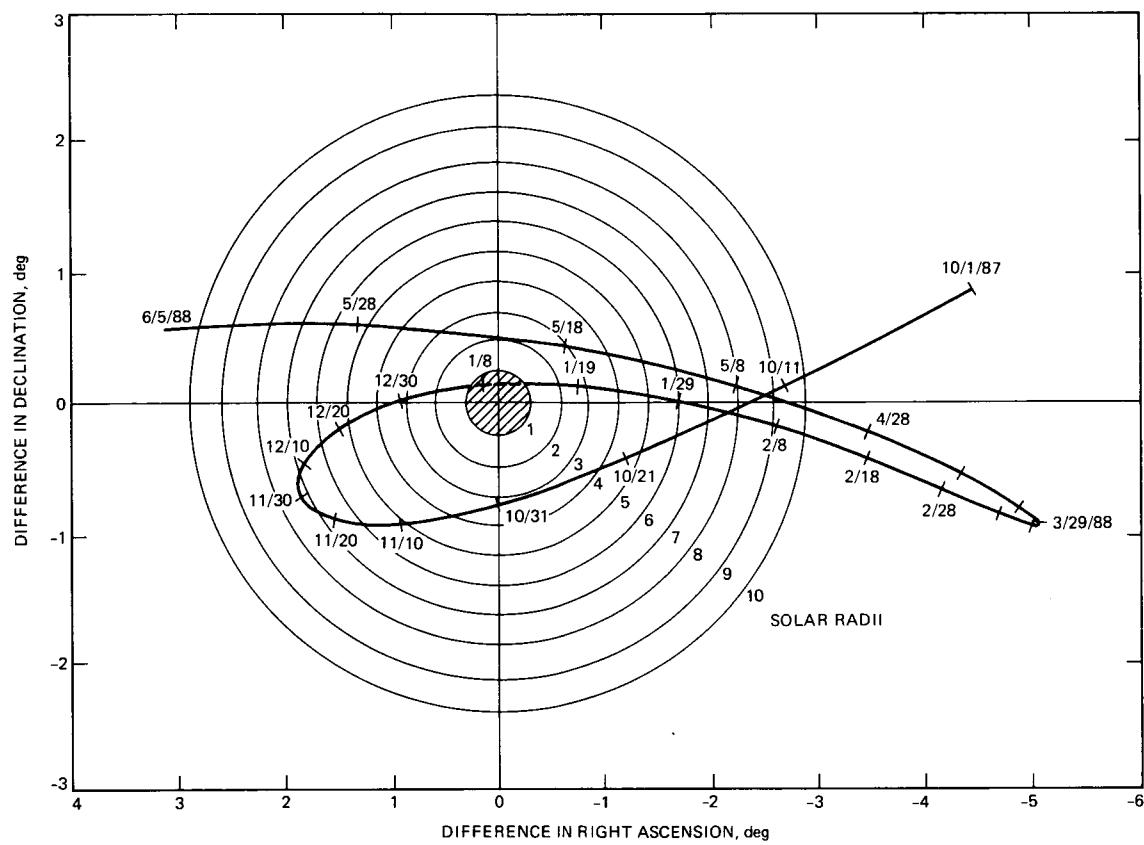


**Fig. 6. Post encounter velocity change inferred from range residuals**

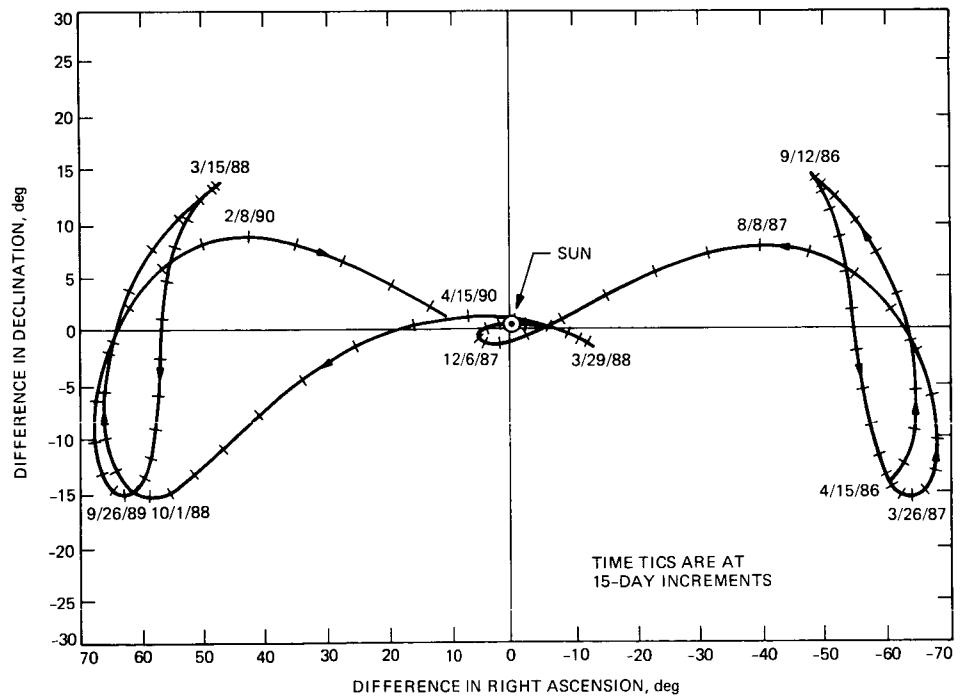


**Fig. 7. DSN Giotto encounter one-way Doppler residuals, 1-second count time**

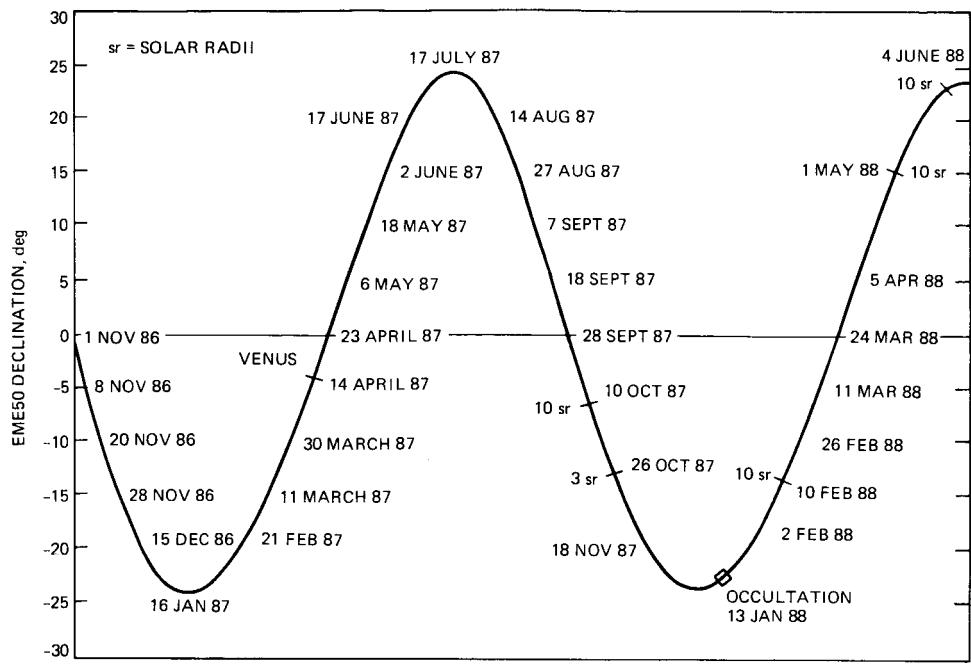
**ORIGINAL PAGE IS  
OF POOR QUALITY**



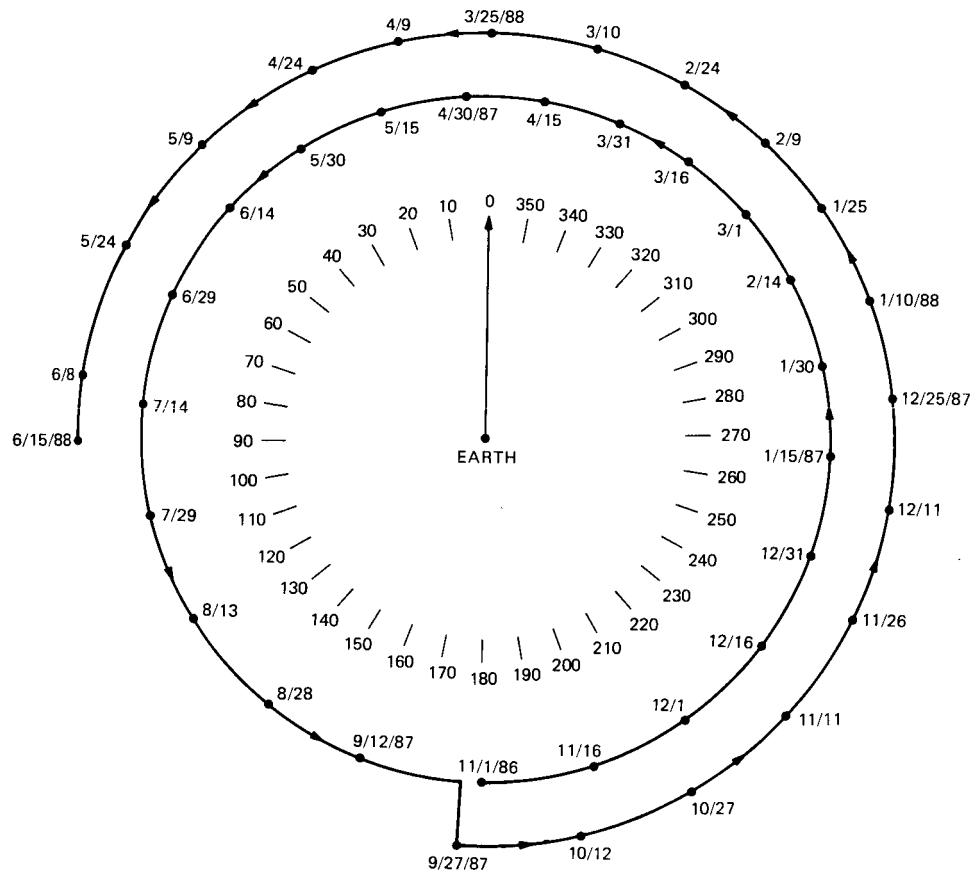
**Fig. 8. Giotto relative to Sun October 1987 to June 1988**



**Fig. 9. Giotto relative to Sun April 1986 to April 1990**



**Fig. 10.** Giotto EME50 declination November 1986 to June 1988



**Fig. 11. Giotto EME50 right ascension November 1986 to June 1988**

# Pathfinder Operations

J. Wilcher and C. Stelzried

TDA Technology Development

S. Finley

Tracking Systems and Application Section

*In 1981, the Inter-Agency Consultative Group (composed of European, Soviet, Japanese and American space agency representatives) conceived the idea of using the two Soviet Vega spacecraft as pathfinders for Giotto since they would arrive at Halley's Comet approximately one week before Giotto. The Vega trajectory data and the Halley camera angle data were combined to improve the comet orbit accuracy. This was used to improve the Giotto fly-by targeting. The DSN performed delta DOR (VLBI) and one-way Doppler measurements of the Vega spacecraft for orbit determination. Although the early start-up phase had many problems, the results during the critical November 30, 1985 to March 4, 1986 operational phase had an overall 95 percent success rate, with 59 successes out of 62 two-station passes.*

## I. Introduction

When the Giotto flyby of the nucleus of Halley's Comet was being planned several years ago, it was realized that a mismatch existed between the requirement to fly close to the nucleus and the knowledge of exactly where the nucleus would be located. In 1981, the Inter-Agency Consultative Group (composed of European, Soviet, Japanese and American space agency representatives) conceived the idea of using the two Soviet Vega spacecraft as pathfinders for Giotto since they were to continue on a trajectory to Halley's Comet after flying by Venus and arriving approximately one week before Giotto.

The concept was to tie the orbit of the comet accurately to an Earth-referenced frame. The trajectory of the Giotto spacecraft, which would also be accurately known from

Earth-based tracking, could be adjusted to pass by the comet at the desired location.

The position of the Vega spacecraft with respect to Halley's Comet could be determined by optical navigation, using on-board imaging systems. The position of the Vega spacecraft with respect to the Earth could be determined with very accurate tracking. The orbits of the Vega spacecraft could be determined using three types of tracking data: Doppler, ranging, and Delta Differential One-way Range (delta DOR), which is based on interferometric techniques and yields the apparent angle from the spacecraft to a reference celestial source (quasars are used) as seen from Earth.

The Soviet's would acquire the first two types of data using their tracking network. It would be the role of DSN network to obtain the delta DOR data. The most critical

period for data acquisition would be that period leading up to the flyby of the comet by the Vega spacecraft. Because of the limited number of delta DOR passes that could be scheduled, seven in February and three in early March, it would be necessary for the DSN to perform flawlessly.

## II. DSS Configuration

The configuration of the DSN 64-m subnet for the Pathfinder Mission was essentially the same as that used for the Venus-Balloon Mission (Ref. 1). The only significant change was the utilization of the DSS Spectrum Processing Subsystem (DSP) in the Very Long Baseline Interferometry (VLBI) mode, which was not utilized for the Venus-Balloon Mission, as the DSP was needed for recovering Telemetry. The configuration is shown in Fig. 1. The received signal was at 1668 MHz (L-band), the same frequency as Venus-Balloon. In order to maximize the reliability of VLBI data return for the Pathfinder Mission, the DSN used redundancy wherever possible. The primary data path was through the VLBI-Radio Science Receiver (MMR) and DSN Spectrum Processing Subsystem. However, to provide a redundant path, the Block IV receiver was configured in an open-loop mode feeding the Block 0 VLBI Subsystem and the Occultation Data Assembly. This provided an alternate path for the data in case of an MMR or DSP failure.

The primary path was the preferred and most used of the paths. The DSP could be configured and controlled through configuration messages routed from the Network Operations and Control Center (NOCC) at JPL via the DSS Monitor and Control Subsystem, thus reducing the DSS operator load. The backup paths required extensive operator intervention to configure and control and were not considered desirable for primary operations.

## III. Mission Operations

Because of the special nature of this mission it was necessary to provide a separate mission operation team and unique procedures to ensure the success of this mission. The role of the special team was to provide the real time and non-real time support to the normal multi-mission operations teams. This effort included the generation of Pathfinder unique Sequence of Events (SOEs), the generation of special antenna pointing predicts, and the real time monitoring of the spacecraft tracking and the data reduction.

### A. Sequence of Events

Because the two Vega spacecraft were under the control of the Soviets and had to be commanded into the 1.67-GHz

delta DOR transmit mode, for each pass it was necessary to construct a very accurate sequence of events. The spacecraft was in a delta DOR transmit mode for only thirty minutes every two hours. Therefore it was necessary to coordinate the spacecraft turn on times such that both spacecraft would be in the delta DOR mode during the short periods of mutual visibility between DSS 63 and DSS 14 and again between DSS 14 and DSS 43.

A typical sequence follows, in Universal Time, Coordinated (UTC):

- 00:00 Vega 2 commanded on by Soviets
- 03:45 Vega 1 commanded on by Soviets
- 09:10 DSS 63 acquires Vega 2, extracts one-way Doppler, verifies spacecraft modes
- 09:45 DSS 63 acquires Vega 1, extracts one-way Doppler, verifies spacecraft modes
- 10:30 DSS 63 reacquires Vega 2
- 11:04 DSS 14/63 Antenna slews to Quasar 3C 273
- 11:10 DSS 14/63 Start DSP recordings
- 11:18 DSS 14/63 Stop DSP recordings, Antennas slew to Vega 1
- 11:23 DSS 14/63 Start DSP recordings
- 11:30 DSS 14/63 Stop DSP recordings, Antennas slew to Vega 2
- 11:35 DSS 14/63 Start DSP recordings
- 11:42 DSS 14/63 Stop DSP recordings, Antennas slew to Quasar DW 1335-12
- 11:47 DSS 14/63 Start DSP recordings
- 11:55 DSS 14/63 Stop DSP recordings, DSS 14 slews to Vega 1, DSS 63 End of Track

DSS 14 continues to track one of the two Vega spacecraft until DSS 43 rises, at which time the delta DOR sequence is repeated.

### B. Predicts

Normally, for delta DOR observations, the spacecraft is tracked using spacecraft predicts and the quasar is tracked using sidereal tracking. The quasar celestial position is used. For the Pathfinder mission, a single program was used to generate the pointing predicts for both the spacecraft and the quasar. This program was a modification of an existing program normally used for radio source tracking. Using a single

program for all pointing eliminated the need for antenna mode switching when changing from spacecraft to quasar.

In addition, these predicts incorporate a basic model of the 1.6 GHz (L-band) pointing offsets required at each 64-m antenna in order to achieve boresight. The offsets are necessary because of physical constraints in mounting the L-band feed. Because of its large physical size, the L-band feedhorn could not be mounted so that its optical axis would lie on the cone described by rotating the Subreflector. Rather, it is mounted between, and slightly outboard of two of the 64-m Tricones. In order to steer the L-band beam to the proper point in the antenna's local celestial sphere, pointing offsets must be applied to the antenna predicts. The required pointing offsets were determined at each 64-m antenna and the offsets incorporated in the antenna pointing predicts program.

### C. Data Processing

Delta DOR data obtained by the participating DSS' were replayed from the DSS' to a central processing facility located in the NOCC. Data from the spacecraft(s) are separated from the data recorded while tracking the quasar, and processed independently (Fig. 2). Both data sets are subjected to the same processing functions: correlation to extract signal phase, data compression to build signal-to-noise ratio, and computation of delay.

**1. QUASAR.** A model of the interferometric phase is developed from the acquisition frequency configuration and values for station location, quasar position, and media delays. A search is made for clock offset between stations and the quasar data are cross-correlated. One second averages of model phase, and sine and cosine correlation coefficients representing residual phase, are computed by the program VCOROP. The program PHASOR then divides the data into sections of about 30 seconds, and outputs model and residual phase for the midpoint of each section.

**2. SPACECRAFT.** A nominal spacecraft trajectory file is used to generate polynomial models for the expected received spacecraft phase at each station. The function Signal Extrac-

tion by Tone Tracking (SEXTNT) uses this information and the program PHASETRK to search in frequency for the offset between the predicted and actual spacecraft signal. A new model polynomial is derived from the data and this data is further processed using the program POSTRK. The output of this process is the spacecraft model phase and the residual phase.

**3. BWS.** The observed delay for both the spacecraft and the quasar is calculated in the BWS program. These delays, along with appropriate time tags, are output to interface with the Orbit Determination Program (ODP).

## IV. Performance

The performance of October and November was not of the quality necessary to support Pathfinder. High data loss during this period was attributed to equipment failures, which is not surprising, since much of the equipment had only recently been installed. This was indicative of the DSP at DSS 63, which had problems with this subsystem throughout this period. There were a few procedural problems, but these soon disappeared as the operations personnel became familiar with the procedures. Probably the most significant item which led to the final success, was the utilization of a predicts program which drove the antennas in a relatively unattended mode. This alone allowed the operators to monitor the operation of the remaining subsystems in a more effective manner. Also, equally important to the success was the diligent monitoring of the operations by the team in the NOCC. They were well prepared with operational work-arounds in case of need. This was demonstrated when the subreflector positioner at DSS 43 failed during a pass and a new pointing offset had to be derived and added to the predict set at DSS 43. The final performance of the DSN for Pathfinder is depicted in Fig. 3 for the critical operations period of November 30, 1985 through March 4, 1986, the end of the Pathfinder operations effort. The overall final performance of 95 percent with 59 successes out of 62 scheduled two-station passes (one spacecraft orientation problem, two DSN problems) for the Pathfinder mission is a credit to all involved personnel.

## Reference

1. Stelzried, C. T., et al., "The Venus Balloon Project," *TDA Progress Report 42-85*, pp. 191-195, Jet Propulsion Laboratory, Pasadena, Calif., May 15, 1986.

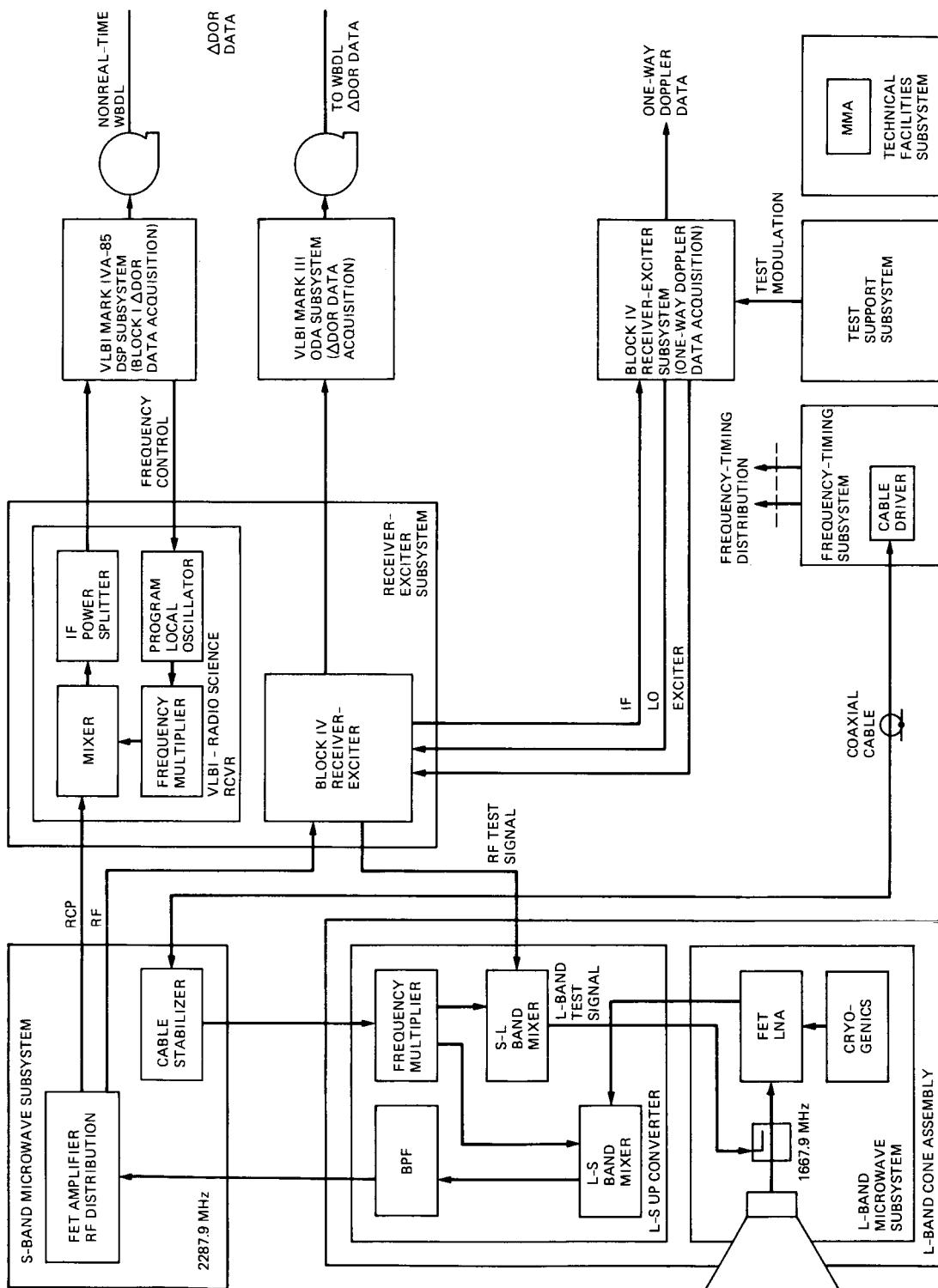
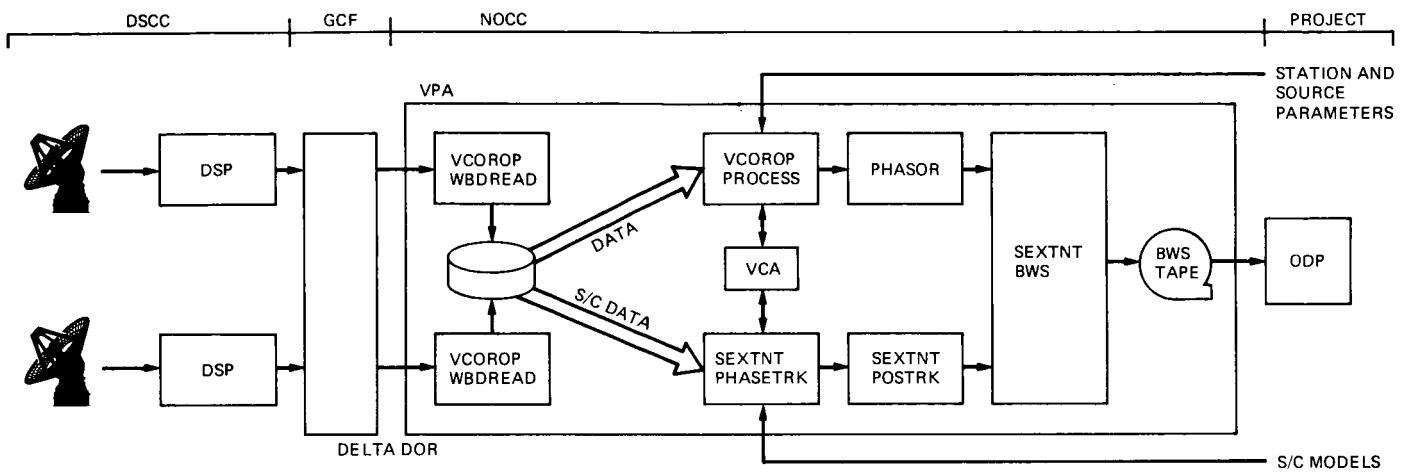
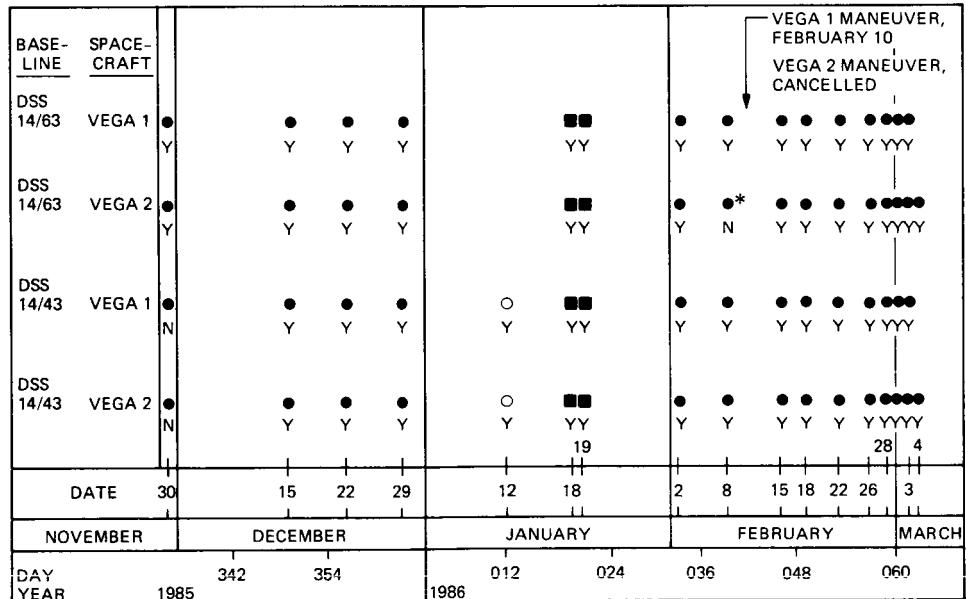


Fig. 1. 1.67-GHz (L-band) configuration block diagram



**Fig. 2. Data processing**



LEGEND: Y = SUCCESS, N = FAILURE ■ TRAINING, CLOCK SYNC ONLY  
○ TRAINING, DSS 43 ONLY  
\* SPACECRAFT ORIENTATION PROBLEM

**Fig. 3. DSN Pathfinder operations performance**

D27-12

N87-15356

TDA Progress Report 42-87

July-September 1986

## VEGA Pathfinder Navigation for Giotto Halley Encounter

J. Ellis

Navigation Systems Section

T. P. McElrath

Federal Electric Corporation

*Results of the VEGA Pathfinder concept which was used to successfully target the European Space Agency's Giotto spacecraft to a 600 km encounter with the comet Halley are presented. Pathfinder was an international cooperative navigation activity involving USSR, European and U.S. space agencies. The final Giotto targeting maneuver was based on a comet location determined from optical data acquired by the earlier arriving Soviet VEGA spacecraft. Inertial pointing angles extracted from optical images of the comet nucleus were combined with a precise estimate of the VEGA encounter orbits determined using VLBI data acquired by NASA's Deep Space Network to predict the location of Halley at Giotto encounter. This article describes the VLBI techniques used to determine the VEGA orbits and shows that the insensitivity of the VLBI data strategy to unmodeled dynamic error sources resulted in estimates of the VEGA orbits with an accuracy of 50 km.*

### I. Introduction

In March 1986, five spacecraft encountered the comet Halley—gathering scientific data and transmitting the first optical images of a comet nucleus. This armada consisted of 2 identical Soviet Spacecraft VEGA-1 and VEGA-2, the Giotto mission launched by the European Space Agency (ESA) and the MS-T5 (Sakigake) and Planet-A (Suisei) missions launched by Japan's Institute of Space Science (ISAS). Encounter conditions for the five probes are summarized below.

### Space Missions to the Comet Halley

Mission (Agency)	Flyby Date, 1986	Distance, km
VEGA-1 (USSR)	6 March	8,900
VEGA-2 (USSR)	9 March	8,000
SUISEI (ISAS)	8 March	151,000
SAKIGAKE (ISAS)	11 March	7,000,000
GIOTTO (ESA)	14 March	600

During the early mission planning phase, an Inter-Agency Consultative Group (IACG) was formed to seek ways of mutual cooperation among the Halley missions. Membership in the Group included delegations from ESA, Intercosmos of the USSR Academy of Science, ISAS and NASA. An out-growth of this effort was the formulation of the Pathfinder concept in which onboard optical data acquired by the earlier arriving VEGA probes as they flew by the comet nucleus were used to improve the comet Halley ephemeris and aid the Giotto terminal navigation (Ref. 1). A schematic diagram of the concept is illustrated in Fig. 1 which depicts the relative flight paths and error ellipses of the VEGA spacecraft, Giotto and the comet at encounter.

The most stringent navigation requirement for the Halley mission set was the 500 km sun-side Giotto encounter. Because the accuracy of the comet location determined from earth-based astrometric data was considerably less than the space-craft location accuracy determined from conventional radiometric tracking data, the comet ephemeris uncertainty was the dominant error in estimating the accuracy of targeting a space-craft to encounter Halley. Estimates of the comet uncertainty ranged from 200 to 1500 km (one sigma). This situation was particularly significant to the Giotto mission which was designed to make direct in situ measurements of the comet's dust and gas composition and transmit television images of the nucleus, therefore requiring both a close flyby and a target accuracy which would ensure a flight path on the sun side.

The Pathfinder concept was a joint NASA, Intercosmos and ESA cooperative effort. The roles of the three agencies were as follows:

- (1) The USSR Space Research Institute (IKI) provided ESA with the inertial camera pointing angle data from the VEGA comet flybys. The 3-axis stabilized VEGA spacecraft were equipped with a TV system which automatically detected and tracked the comet during the encounter phase. Inertial comet pointing angles (right ascension and declination) were extracted from the TV system for a two hour period at each encounter. To use this information, it was necessary to determine the location of the VEGA probes at the time of comet encounter (Refs. 2-4).
- (2) NASA's role was to track the VEGA probes using Very Long Baseline Interferometry (VLBI) techniques and precisely estimate the VEGA encounter orbit. It was recognized that estimates using conventional two-way range and doppler, as employed by IKI, could not achieve the level of accuracy required for Pathfinder. However, this level of accuracy could be met using VLBI data acquired by NASA's Deep Space Network (DSN) (Refs. 5, 6). NASA's Jet Propulsion Laboratory

(JPL) had pioneered the use of VLBI for deep space navigation of the Voyager probes and had applied the technique to determine the Venus relative orbits of the VEGA probes for the Venus Balloon Experiment in June 1985.

- (3) The final Giotto target maneuver, executed two days before the Halley encounter, was based on an updated comet ephemeris determined by ESA using the IKI Pathfinder camera information and the NASA VLBI orbit estimate combined with earth-based astrometric observations.

Pathfinder required a stringent timeline for the exchange of information—especially to ensure that the results from the March 9th VEGA-2 encounter were available in sufficient time for the March 12th Giotto maneuver. A technical team was formed to implement this concept consisting of members from ESA, IKI and JPL. The team prepared a series of documents (Refs. 7-9) describing the project requirements, data interfaces and operations schedule. Test activities were defined and executed during the Venus-Halley cruise phase to verify the compatibility of models, test the data interfaces and simulate operations during the encounter phase.

The critical technical activities of the Pathfinder project occurred during the 48 hour period following each VEGA Halley encounter. The DSN concluded its tracking of the VEGA spacecraft on March 4th, and VEGA orbit solutions using the DSN VLBI data were determined independently by JPL and IKI, compared and transmitted to ESA. After each VEGA flyby, camera images were transmitted to IKI and direction angles from the VEGA probe to the comet nucleus were computed and relayed to ESA within 24 hours. The position of the comet nucleus as computed by both IKI and ESA using the comet angle data and VEGA VLBI orbits agreed to within 30 km. Pathfinder reached a successful conclusion on March 11 when ESA and IKI agreed on a final comet nucleus position and estimated the Giotto-Halley target line accuracy to be 40 km. Based on this information, the Giotto Science Working Team recommended targeting Giotto to an encounter at 500 km + one sigma (40 km) from the nucleus and 20 degrees below the comet-sun line. The one sigma error was added to increase the probability of a flyby distance greater than 500 km. A velocity change of 2.5 meters/sec was executed by ESA on March 12th to achieve the desired target point.

Results from the VEGA Pathfinder were not only critical for targeting the Giotto encounter but provided the basis for improving the observational model used to process the ground-based International Halley Watch (IHW) data. The discrepancy between the IHW and VEGA-1 Pathfinder comet Halley ephemeris was 248 km in the Giotto target plane. Based on

this difference, Giotto maneuver planning was delayed until the VEGA-2 Pathfinder results were available. The VEGA-2 results essentially confirmed the VEGA-1 Pathfinder estimate. The final Giotto target maneuver was based on a comet Halley ephemeris update using the combined Pathfinder data. Upon subsequent analysis, the IHW post-perihelion astrometric data was revised to account for a significant offset between the comet center of light and center of mass. Once this effect was modeled in the IHW data, the Pathfinder and IHW ephemerides agreed to within 50 km in the Giotto target plane.

This article describes the results of the VEGA Pathfinder effort. The principal focus is on the VLBI techniques and the results of the VLBI orbit determination.

## II. DSN VLBI Tracking

### A. The VLBI Observable

The VLBI technique uses two widely separated tracking sites to simultaneously receive the wideband signal broadcast by the VEGA probes. As shown in Fig. 2, cross correlation of the signal furnishes a precise measure of the difference in arrival time of the signal at the two stations—which determines the angle between the interstation baseline and the source. By alternately tracking the VEGA probes and an angularly nearby extragalactic radio source (EGRS) or quasar, whose location is known, a doubly differenced measurement is formed in which common errors are canceled. This data type is called delta Differential One Way Range or  $\Delta$ DOR. Differencing the EGRS and VEGA signals cancels errors due to clock synchronization, transmission media and platform parameter uncertainty, with the degree of cancellation dependent on the probe-source separation. Two nearly orthogonal baselines are required to resolve the geocentric right ascension and declination of the source. For DSN VLBI, this situation is satisfied by the Goldstone-Canberra and Goldstone-Madrid baselines.

### B. DSN Implementation and Operations Support

The VEGA mission consisted of a Venus flyby phase and a Halley encounter phase. In mid June 1985, each spacecraft encountered Venus and successfully released an entry probe and a wind-measuring balloon into the Venus atmosphere. Post Venus encounter maneuvers were executed following the flyby to target the probes to a March comet encounter. The VEGA spacecraft and the balloons carried a stable crystal oscillator which was used as a reference for transmitting an L-band signal at 1.668 GHz. The L-band frequency was selected specifically for the Venus Balloon Experiment to be compatible with the reception capabilities of an international network of 20 radio observatories. Conventional (two-way) tracking and commanding of the VEGA spacecraft used a C-band frequency at approximately 6 GHz.

The balloon experiment was a cooperative venture of the Soviet Union and France. The international network, which included the DSN 64 meter subnet, was organized by the French space agency Centre National D'Etudes Spatiales (CNES) to receive the one way L-band signal broadcast by both the VEGA probes and the balloons during the 48 hour balloon lifetime. The DSN modified the 64 meter antenna subnet to receive this L-band signal. An L-band microwave feed horn subsystem was mounted on the 64 meter antennas and a low noise amplifier and frequency upconverter were configured to convert the L-band spectrum to a DSN compatible S-band signal for input to the S-band microwave subsystem. The DSN L-band capability was used for the Venus Balloon Experiment and for the Pathfinder operations.

The VEGA L-band signal consisted of either a pure carrier or two subcarriers separated by 6.5 MHz. The two subcarrier tones were transmitted for half hour periods every two hours. This signal was commanded on and off by a Soviet tracking station. Consequently, for Pathfinder operations transmission times were explicitly coordinated between IKI and JPL to ensure that the tones were transmitted during the interval that the two DSN stations simultaneously viewed the spacecraft.

A program to construct an L-band catalog of natural radio sources to be used for the VEGA VLBI observations was organized by the DSN. This consisted of selecting known sources from the S/X-band VLBI catalog that were in the vicinity of the VEGA flight path and validating that the source structure and correlated flux density were suitable at L-band. The locations of the sources relative to the VEGA orbits are plotted in Fig. 3.

## III. VEGA VLBI Orbit Determination

### A. Data Processing Strategy

Both JPL and IKI independently determined the VEGA flyby orbits using the DSN  $\Delta$ DOR data acquired during the encounter phase. VLBI observations from Goldstone-Canberra and Goldstone-Madrid baselines for both spacecraft were acquired weekly in December 1985 and approximately twice per week starting in February 1986. This consisted of 3 observing sessions in December, 7 in February and 3 in March. The final VLBI observations for VEGA-1 were acquired on March 3 (Encounter-3 days) and for VEGA-2 on March 4 (E-5 days). A typical VLBI observation session from a single baseline consisted of a EGRS, VEGA-1, VEGA-2, and EGRS scan sequence with 7 minute scans for each spacecraft and 7 minute scans for the natural radio sources.

IKI orbit estimates were determined by fitting the IKI two-way range and doppler data and DSN  $\Delta$ DOR observations. Because the IKI tracking philosophy consisted of acquiring

short (20 minute) passes of two way data twice per week from two tracking sites, the estimate using only two-way data required long arcs to achieve a reasonable degree of accuracy and did not result in a consistent estimate. For this reason, IKI chose to use the  $\Delta$ DOR solutions for their own encounter planning. The JPL orbit estimates were based on a combination of  $\Delta$ DOR measurements and geocentric range and range-rate measurements. The latter were constructed from IKI state vector information and two-way range and doppler residuals with respect to this state. Typically, one pair of geocentric observations were included biweekly.

The criterion for formulating a data processing strategy to fit the data was to select a strategy which would be insensitive to unmodeled dynamic error sources. In particular, JPL was not always informed about spacecraft events which could have had an effect on the orbit estimates. Soviet experts indicated that velocity variations could be expected due to attitude control maneuvers of 1 cm/sec over 1 day and that the solar pressure constant could be in error by 15 to 20%. Also, IKI was having difficulty converging to a consistent solution with long arcs of two-way data and the values of the solar pressure constant estimated from such arcs differed considerably from the theoretical value. IKI attributed the solution inconsistency to unmodeled nongravitational effects such as frequent attitude maneuvers. In addition, for VEGA-1 a final Halley target maneuver of approximately 18 meters/sec was executed February 10 (Encounter-24 days) with an execution error estimate of 1 meter/sec. An a priori maneuver estimate of the delta-V's was provided by IKI. No additional encounter maneuver was required for VEGA-2.

Because of the concern with unmodeled dynamic error sources, JPL preferred to rely on short arc solutions which tend to be less sensitive to such errors. For short arc solutions the accuracy is dominated by the measurement error while for long arc solutions dynamic errors can be expected to dominate. The JPL solution strategy was based on selecting the arc length and estimated parameter set which simultaneously yielded the best fit to the observations and minimized the consider covariance. Parameters which influenced the uncertainty of the estimate but could not be adequately determined by the filter were included or considered in computing the statistics of the estimated parameters.

For the maneuver-free VEGA-2 arc, the baseline strategy was to fit the February-March data arc (E-35 to E-5 days) and estimate state only. Solar pressure acceleration errors were considered in computing the statistics of this solution. Because of the VEGA-1 maneuver at E-24 days, the VEGA-1 solution was based on fitting data starting in December (E-91 to E-3 days) and estimating state, solar pressure and maneuver components.

All solutions assumed a  $\Delta$ DOR data weight of 1 meter, geocentric range of 10 km and range-rate of 0.1 meters/sec. The  $\Delta$ DOR error model was comprised of systematic and random error sources. The random errors included station oscillator errors, dispersive instrumental phases errors and SNR for the spacecraft and quasar signals. Tropospheric and ionospheric errors were combined into individual bias errors for each baseline. The natural radio source position error consisted of a frame tie error which characterized the uncertainty of the quasar catalog reference frame with respect to the planetary FK-4 frame and a relative error which described the uncertainty of quasar locations within the radio source reference frame. The 250 nanoradian frame-tie uncertainty was based on an estimate of the error derived from the VEGA Venus flyby solutions for the Venus Balloon Experiment. Quasar position errors and  $\Delta$ DOR bias errors were treated as consider parameters in computing the statistics of the solutions. Table 1 summarizes the filter model assumptions.

## B. Solution Convergence

Solution convergence was evaluated by plotting the B-plane estimates as a function of the data termination time. Figure 4 displays the sequence of VEGA-1 and VEGA-2 B-plane solutions with respect to the final converged solution. The one sigma error ellipses are also plotted for each solution. Only the postmaneuver solutions are plotted for VEGA-1 with tracking data arcs terminating at E-3 days, E-8 days and E-15 days. The comparatively large uncertainty at E-15 days reflects the uncertainty in the estimates of the maneuver at E-24 days. VEGA-2 solutions based on data from E-35 days to E-5 days, E-12 days and E-19 days are plotted in Fig. 4(b). Rapid convergence to the final solution is attained in each case. The final solution uncertainty, as will be shown later, is dominated by the frame-tie error.

## C. JPL-IKI Final Orbit Determination Results

Solutions computed using the baseline strategy were evaluated by comparing them with estimates derived using alternative procedures. Figure 5 displays the JPL and IKI comet-relative B-Plane solutions and corresponding one sigma consider error ellipses with the origin of each plot at the JPL baseline solution. The B-Plane values are computed with respect to JPL Halley comet ephemeris (DE118) HL39. The effect of comet ephemeris uncertainty is not included in the statistics. The nominal long-arc VEGA-1 solution is compared with the IKI solution and a shorter post-maneuver arc (E-19 to E-3 days) case. Solar pressure acceleration errors were considered for all short arc cases. Because of the maneuver, the VEGA-1 estimates were relatively insensitive to the choice of strategy with the JPL and IKI solutions agreeing to within 15 km in B.R and 5 km in B.T. The VEGA-1 encounter maneuver was determined with an accuracy of

0.03 meters/sec and the solar pressure with an accuracy of 10% of the nominal value. For VEGA-2, the nominal short arc solution is compared with the IKI solution and a long arc case (E-95 to E-5 days). Although the uncertainty of the estimate decreased for the long arc VEGA-2 case, the ability to fit the  $\Delta$ DOR data degraded. The baseline JPL and IKI VEGA-2 solutions differed by less than 5 km. However, the long arc solution, which was more sensitive to unmodeled dynamic errors, differed from the short arc case by 37 km in B.R and 6 km in B.T.

The effects of the individual error sources on the encounter statistics are plotted in Fig. 6. The errors are expressed in a geocentric reference frame with one axis along the earth-spacecraft direction and the other two orthogonal axes in the right ascension and declination directions. The two angular components are directly determined by the  $\Delta$ DOR data and the geocentric range by the quasi-geocentric range data. The limiting error is due to the 250 nanoradian radio-optical frame tie uncertainty.

The  $\Delta$ DOR residuals for the JPL solutions are displayed in Fig. 7. Residuals for both solutions have a one-sigma standard deviation of 0.6 m. The residuals were corrected for tropospheric but not ionospheric calibration errors. Tropospheric corrections were based on a standard wet and dry component model for the DSN stations. Attempts to correct the data for ionospheric using Faraday rotation data yielded inconsistent results—which may reflect on the quality of the corrections derived from the Faraday data. Including the ionospheric corrections decreased the VEGA-1 residual errors by 30%, but increased VEGA-2 residuals by 20%. The change to the baseline solutions was insignificant.

The standard deviation of the geocentric range and range-rate residuals were 3.1 km and 0.55 cm/sec for VEGA-1 and 1.6 km and 0.3 cm/sec for VEGA-2.

#### **D. Attitude Control Maneuver Sensitivity**

Throughout the approach phase frequent attitude maneuvers were executed to maintain the spacecraft orientation. The sensitivity of the baseline orbit estimates to such effects was investigated by assuming that the dynamics were corrupted by gaussian white noise accelerations of  $1.0 \times 10^{-10}$  km/sec<sup>2</sup> acting along each axis of the spacecraft. A batch sequential filter and smoother procedure was used to estimate the stochastic accelerations assuming a 7 day batch size. The results are summarized in Table 2.

The net effect of including the stochastic accelerations was to improve the fit of the  $\Delta$ DOR data without significantly affecting the solution. The VEGA-1 solution is less sensitive to stochastic acceleration effects due to the maneuver. The

relative insensitivity of the solutions to stochastic accelerations demonstrates the strength of the information content of the  $\Delta$ DOR data. With conventional two way doppler and range a long arc would have been required to determine the orbit. The use of  $\Delta$ DOR enabled the solution to be determined by a short arc which is less sensitive to unmodeled acceleration effects.

#### **IV. Pathfinder Results**

The VEGA VLBI orbit determination was just one element of the Pathfinder concept. This information was combined with inertial camera pointing angle data to determine the comet ephemeris at the time of Giotto encounter. Figure 8 reconstructs the situation prior to the final Giotto trajectory correction maneuver (TCM) on March 12th. The comet relative B-Plane locations of Giotto (prior to this TCM) are plotted based on the following assumptions:

- (1) JPL Halley Comet Ephemeris (DE118)HL39 which included IHW postperihelion astrometric data to February 17. This essentially was the ephemeris used at JPL prior to any Pathfinder results and was similar to the ephemeris used at ESA.
- (2) The comet ephemeris as determined from the VEGA-1 Pathfinder results (PF01).
- (3) The comet ephemeris resulting from the combined VEGA-1 and VEGA-2 pathfinder results (PF03). This was designated as the official combined VEGA-1 and VEGA-2 solution. The combined solution (PF05) which was used for the encounter TCM is also plotted.
- (4) The JPL Halley Comet Ephemeris (DE118)HL45 which was constructed after the VEGA-1 Pathfinder results were available and accounted for a comet center-of-light center-of-mass offset of 1100 km at 1 AU.

As can be seen from this figure, the VEGA-1 Pathfinder result differs from the pre-Pathfinder JPL ephemeris by 229 km. ESA reported a difference of 248 km from their 'pre-Pathfinder' ephemeris. Based on this difference, Giotto maneuver planning was delayed until the VEGA-2 results were available. The combined VEGA-1 and VEGA-2 solution (PF03) was within 50 km of the VEGA-1 results with the largest difference observed in the B.R component. When the JPL ephemeris was subsequently updated (HL45) to include a center-of-light center-of-mass offset for postperihelion observations, the difference between the JPL and PF03 target plane state was reduced to 27 km.

Two days prior to the Giotto Halley encounter, the final decision was made to target Giotto to a sunward side encoun-

ter at 500 km + one sigma (40 km) from the nucleus and 20 degrees below the sun line. The 40 km uncertainty is a formal predicted uncertainty, as determined by ESA (Ref. 10), which includes the Pathfinder determination of the comet location and spacecraft errors. The predicted B-Plane aim point was  $B.T = -507.4$  km and  $B.R = -184.7$  km. Based on pre- and postencounter tracking data and the best estimate of the encounter parameters was  $B.T = -545 \pm 30$  km (1-sigma) and  $B.R = -275 \pm 80$  km.

The Pathfinder results not only played a critical role in the final Giotto navigation but also influenced the development of the comet observation model. The ability to fit earth-based observations of the comet Halley collected after the comet's perihelion passage on February 9th, 1986 degraded significantly. A systematic bias of 2 arcseconds was observed in the astrometric measurements collected following perihelion which translated into a position error of 1500 km. This was believed to be largely due to the center of light, center of mass difference caused by the increased activity of the comet and the short term fluctuations in this activity. Since this bias lies along the sun-comet vector, its effect is not separable from the comet nongravitational accelerations and consequently is nonobservable. The consistency of the Pathfinder solutions confirmed the need to model a bias to account for the light

shift and aided in the development of an empirical bias model for processing the ground-based observations.

## V. Conclusions

Based on the results of the error analysis and our observations of the solution consistency, the maximum error in the Earth relative determination of the VEGA encounter states was 50 km. In terms of comet relative B-plane components the critical  $B.T$  direction was determined to better than 30 km. Without the DSN VLBI data, it is unlikely that short arc solutions could have been employed to achieve this level of accuracy.

The Pathfinder encounter essentially provided the comet information required for targeting Giotto to a close sun-side comet encounter. The estimate of the comet location at the time of Giotto encounter was improved and the uncertainty of this estimate was significantly reduced. A premaneuver decision was reached by the Giotto Science Working Team to target Giotto to a 500 km plus one sigma encounter distance. The ESA determination of this one sigma uncertainty using the Pathfinder data was approximately 40 km. A preliminary assessment of the Giotto camera data showed that Giotto achieved a comet encounter at a distance of between 580 and 605 km.

## Acknowledgment

The authors wish to acknowledge the support provided by Dr. J. F. Jordan and Dr. C. T. Stelzried for their invaluable leadership in guiding the successful implementation and operations of the Pathfinder effort and the support provided by S. G. Finley, R. D. Shaffer and J. H. Wilcher for their heroic efforts in planning and ensuring the successful acquisition of the VEGA VLBI data.

## References

1. Reinhard, R., "Exploration of Halley's Comet from Space: The Inter-Agency Consultative Group (IACG) and its Associated Working Groups," *ESA Bulletin No. 38*, July 1984.
2. El'yasberg, P., et al., "A Priori Prediction Accuracy of Comet Halley Position Using Pathfinder Concept," Presented at 2nd Pathfinder Workshop, Institute of Space Research, Moscow, USSR, June 1983.
3. Hechler, F. L., et al., "Incorporation of VEGA Pointing Angles and Spacecraft Ephemeris into the GIOTTO Terminal Navigation Process," *Giotto Study Note No. 37*, ESOC, ESA, Darmstadt, May 1983.
4. de Broeck, P., "Error Characteristics for the VEGA Observations of the Comet Halley and Reviewed Obtainable Giotto Targeting Accuracy," *Giotto Study Note No. 55*, ESOC, ESA, Darmstadt, August 1985.
5. Campbell, J. K., Ellis, J., and Jordan, J. F., "Pathfinder: A Technique for Improving the Targeting Accuracy of Giotto," *AAS Paper 83-417*, Lake Placid, N.Y., August 1983.
6. Ellis, J., Deep Space Navigation with Noncoherent Tracking Data, *TDA Progress Report 42-47*, pp. 1-12, Jet Propulsion Laboratory, Pasadena, Calif., April-June 1983.
7. INTERCOSMOS-ESA-NASA, "PTP Pathfinder Technical Project," IACG-PF-PTP/1-01 Pathfinder Documentation, ESOC, Darmstadt, October 1984.
8. INTERCOSMOS-ESA-NASA, "Pathfinder Operations Schedule," IACG-PF-OS-01, Pathfinder Documentation, ESOC, Darmstadt, October 1985.
9. INTERCOSMOS-ESA-NASA, "Pathfinder Systems Test Plan," IACG-PF-STP-01, Pathfinder Documentation, ESOC, Darmstadt, September 1985.
10. Morley, T.A., "Comet Halley Orbit Determination Vol. IV to Encounter, March 1986," *Giotto Flight Dynamics Report No. 1 IV*, ESOC, ESA, Darmstadt, May 1986.

**Table 1. Nominal filter error model assumptions**

Parameter	A Priori Standard Deviation	
Estimated Parameters	VEGA-1	VEGA-2
Position	$1 \times 10^5$ km	$1 \times 10^5$ km
Velocity	10 km/s	10 km/s
Maneuver	1 m/s	—
Solar Pressure Acceleration	$2 \times 10^{-10}$ km/s <sup>2</sup>	—
Consider Parameters		
Solar Pressure Acceleration	—	15% ( $0.6 \times 10^{-11}$ km/s <sup>2</sup> )
EGRS-Radio Frame Tie	250 nrad	250 nrad
EGRS Relative Location Error	50 nrad	50 nrad
ΔDOR Bias Goldstone-Canberra	0.28 m	0.28 m
ΔDOR Bias Goldstone-Madrid	0.60 m	0.60 m

**Table 2. Effect of stochastic accelerations**

ΔDOR 1-Sigma Error, m	Change to Baseline Solution		
	B.R., km	B.T., km	S, km
VEGA-1	0.3	-8.4	5.6
VEGA-2	0.4	-24.5	2.0

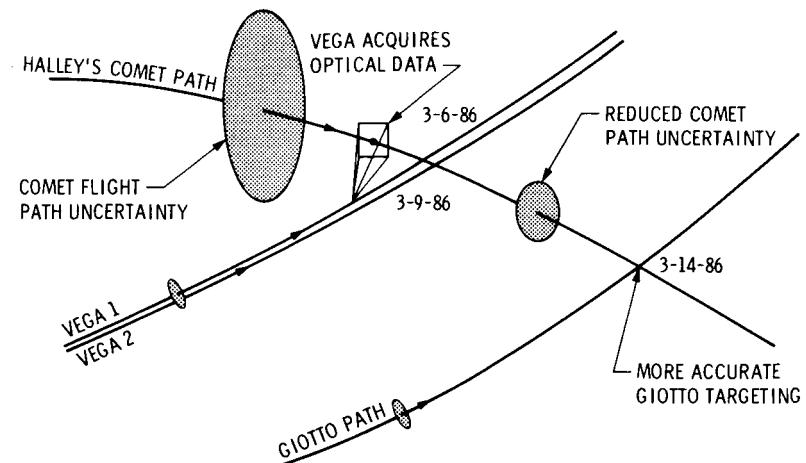


Fig. 1. Schematic diagram of Pathfinder concept

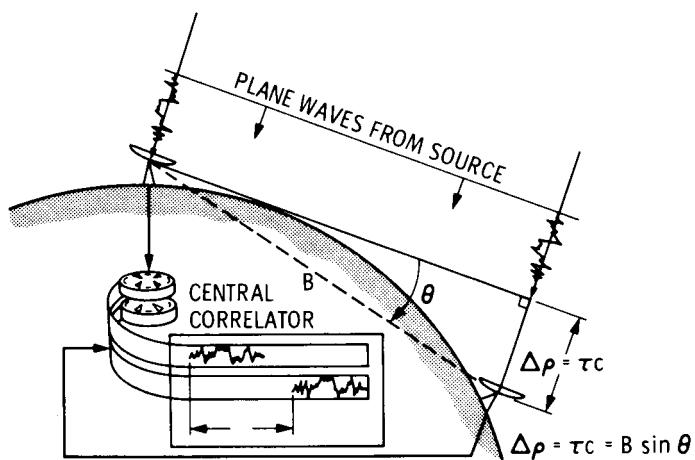


Fig. 2. VLBI observable

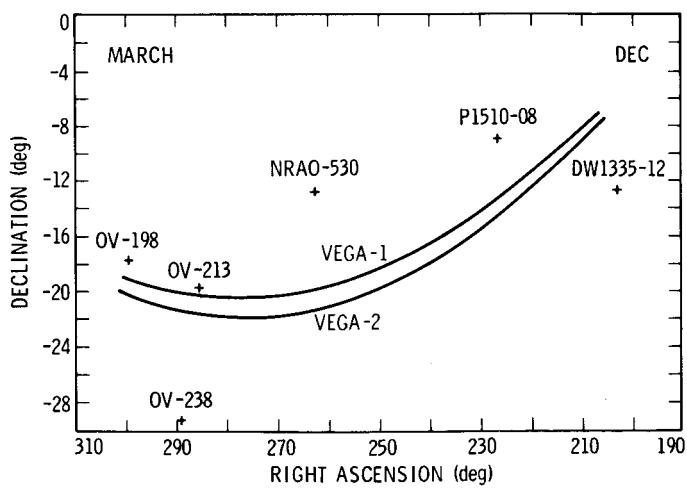
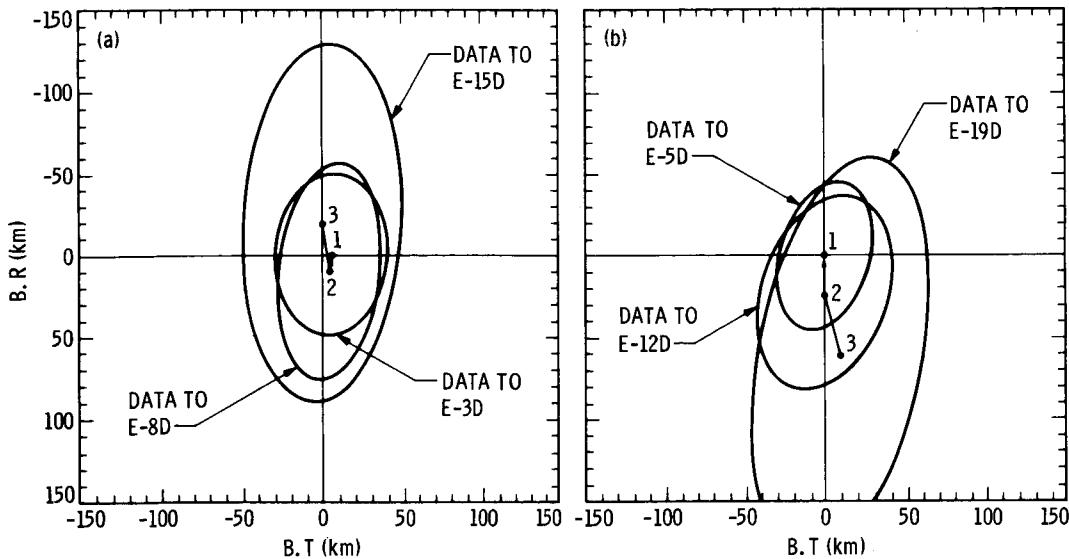
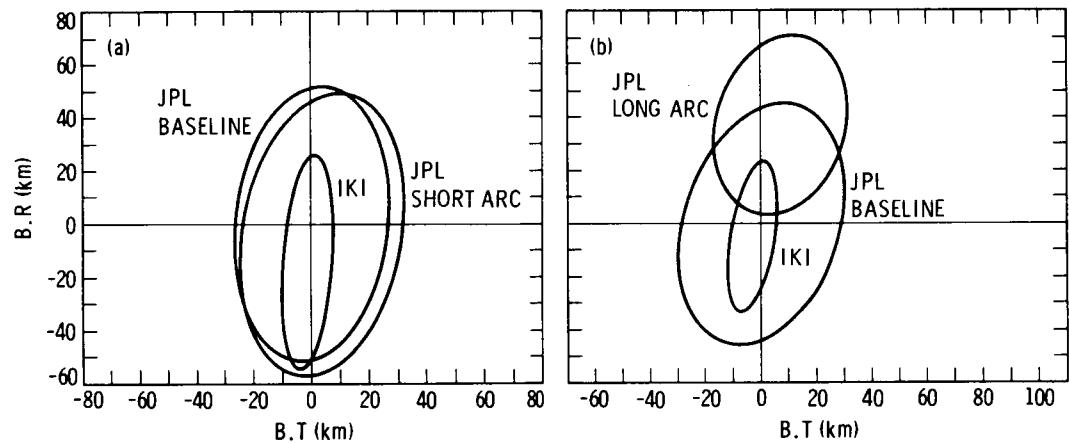


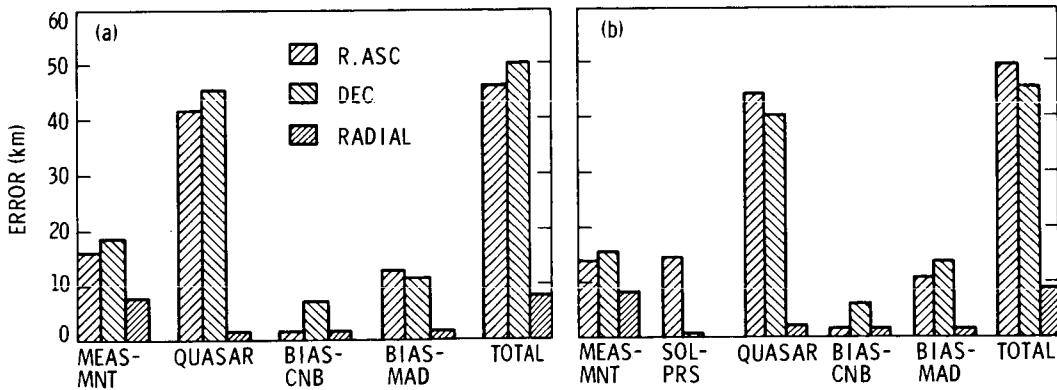
Fig. 3. VEGA L-band radio sources



**Fig. 4. B-Plane convergence: (a) VEGA-1, (b) VEGA-2**



**Fig. 5. Final B-Plane results: (a) VEGA-1, (b) VEGA-2**



**Fig. 6. Error sources: (a) VEGA-1, (b) VEGA-2**

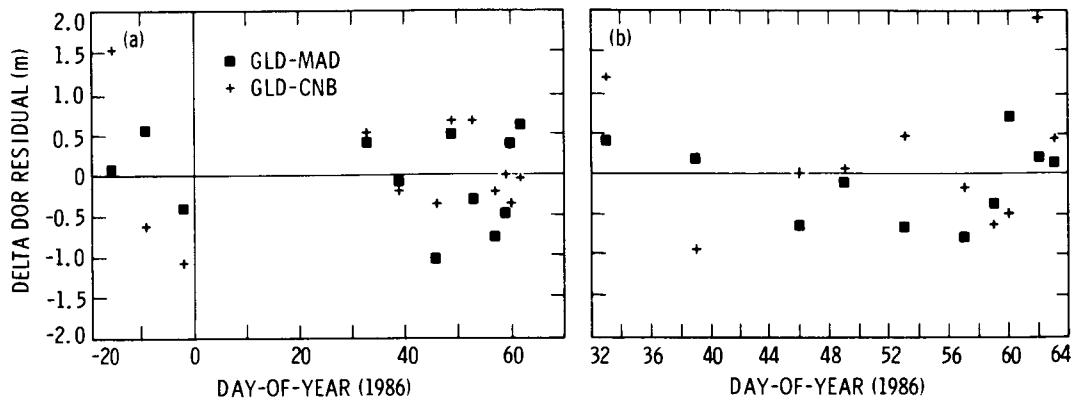


Fig. 7. The  $\Delta$ DOR residuals: (a) VEGA-1, (b) VEGA-1

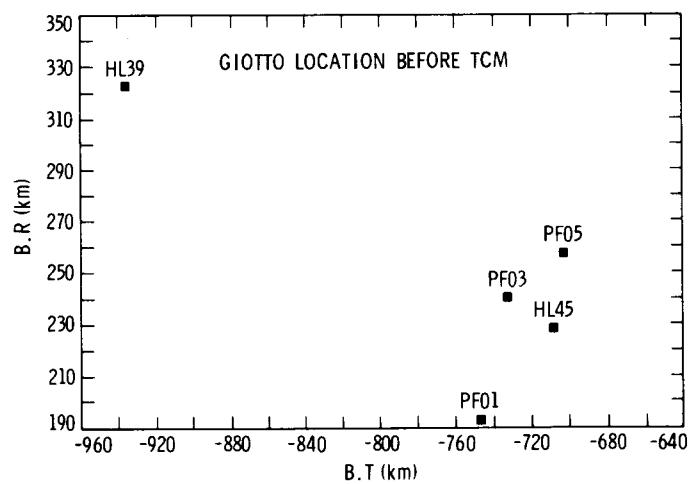


Fig. 8. Giotto B-Plane

# Selection of Radio Sources for Venus Balloon-Pathfinder ΔDOR Navigation at 1.7 GHz

K. M. Liewer<sup>1</sup>  
TDA Engineering Office

*In order to increase the success rate of the ΔDOR VLBI navigational support for the French-Soviet Venus Balloon and Halley Pathfinder projects, forty-four extragalactic radio sources were observed in advance of these projects to determine which were suitable for use as reference sources. Of these forty-four radio sources, taken from the existing JPL radio source catalogue, thirty-six were determined to be of sufficient strength for use in ΔDOR VLBI navigation.*

## I. Introduction

The technique of Very Long Baseline Interferometry (VLBI) (Refs. 1-3) is used to achieve the navigational accuracy required by many projects (Refs. 4, 5). The particular method used for navigation of the Soviet spacecraft VEGA is the ΔDOR (delta-Differential One-way Range) technique. This technique estimates the angular position of the spacecraft on the plane of the sky relative to a reference radio source. Delta-DOR uses the difference in arrival time of a spacecraft signal between two stations (equivalent to a differential one way range measurement) and differences this result with a similar measurement made on a nearby extragalactic radio source (EGRS). The differencing of the spacecraft delay and the EGRS delay results in the partial cancellation of many error sources that are common to the two measurements. This differencing also means the spacecraft position is only determined relative to that of the EGRS. As a result, the posi-

tions of the EGRS must be well known if the ΔDOR measurements are to be useful for navigation.

The degree to which the common error sources are canceled in ΔDOR measurements increases as the angular separation of the spacecraft and reference radio source decreases. It is therefore desirable to have a distribution of EGRS on the sky with sufficient density so that the angular distance between a spacecraft and one of the EGRS is 10 degrees or less. The Deep Space Network currently observes more than 130 extragalactic radio sources in order to obtain a catalogue of source positions with sufficient accuracy and coverage to meet navigational requirements. The accuracy of this catalogue is currently about 50 nanoradians (10 milliarcseconds).

The catalogue available prior to the VEGA mission was derived from observations made at 2.3 GHz (S band) and 8.4 GHz (X band) with most of the weight from the 8.4 GHz observations. The transmissions from the VEGA spacecraft, however, were at 1.7 GHz (L band). This presented two possible problems: (1) The source position at 1.7 GHz could

---

<sup>1</sup>The work described in this article was performed when the author was a member of the Tracking Systems and Applications Section.

differ from that determined from the 2.3/8.4 GHz observations; and (2) the correlated flux density, which is the strength of a source measured with VLBI, was not known at 1.7 GHz for any of the sources in the catalogue.

Because of the small amount of time available for making new observations and the lack of dual frequency observations to correct for charged particle effects, 1.7 GHz positions with errors at the 50 nanoradian level could not be produced. Inspection of the radio maps of several radio sources made near 1.7 and 8.4 GHz indicated that the positions determined using these frequencies would differ by less than 10 nanoradians for most sources (J. S. Ulvestad, private communication, 1985). The decision was therefore made to adopt the positions from the existing 2.3/8.4 GHz catalogue.

Because the transmission schedules of the Russian space-craft were such that there were relatively few opportunities for making  $\Delta$ DOR observations, it was very important that the success rate for these observations be as high as possible. To be assured that the observations did not fail due to lack of sufficient signal from the EGRS, measurements of the L band correlated flux density of potential navigation sources were necessary prior to the  $\Delta$ DOR observations.

## II. Description of Observations

The BLK 0 VLBI system was used to generate the VLBI data. This system samples a 1.8 MHz bandpass and allows measurements to be made on sources that are too weak to be observed with the BLK 1 (250 kHz) system which is used for the actual  $\Delta$ DOR observations. Observations were made in three time-multiplexed channels within the 18-MHz bandpass of the 1.7-GHz receiver. The center frequencies of the channels were 1659.1, 1671.1 and 1676.1 MHz.

The processing of the VLBI data yields a correlated amplitude,  $CA$ , which is given by

$$CA = \frac{SC(\eta_1\eta_2A_1A_2)^{1/2} \times 10^{-26}}{2k(T_1T_2)^{1/2}} \quad (1)$$

where  $S$  is the correlated flux density in Janskys,  $\eta_1$  and  $\eta_2$  are the aperture efficiencies,  $A_1$  and  $A_2$  are the antenna areas,  $k$  is Boltzmann's constant,  $T_1$  and  $T_2$  are the total system temperatures and  $C$  is a constant ( $C = 0.807$ ) which accounts for several effects in the VLBI signal processing. Measurements of the system noise temperature at each antenna and the antenna aperture efficiencies were required in order to calculate the correlated flux densities.

The system temperatures required for converting correlated amplitudes to flux densities were obtained from calibrated strip charts. The efficiency of each antenna is required to get absolute flux from temperature measurements. The efficiency of the 1.7 GHz system at DSS 14 was estimated using radio sources of known flux density in an experiment conducted prior to the VLBI observations.

Observations began in May of 1985 shortly after the 1.7-GHz system on the 64-m antennas became operational. A total of six experiments was performed: Three on the California-Spain baseline and three on the California-Australia baseline. The last observations were conducted in October of 1985. The actual observation time of each experiment was 4 hours. The correlated flux density can have a complicated dependence both on the interferometric hour angle (IHA) and the baseline vector due to spatial structure in the EGRS. The IHA is the difference between the Right Ascension of the perpendicular to the equatorial projection of the baseline vector and the Right Ascension of the source. An attempt was made to observe each source on both baselines and over as wide a range of IHA as possible since it was not known what the IHA would be during  $\Delta$ DOR observations.

The list of sources to be observed was generated from the existing source catalogue by first selecting all sources that were within 30 deg of the VEGA trajectory (for the period May 1985 to June 1986) and then eliminating those weak sources that were unlikely to be used since they were far from the trajectory or near very strong sources. Because of the criticality of the  $\Delta$ DOR observations around the region of encounter, all sources in that region were observed regardless of estimated source strength.

## III. Results

The total flux density measurements of radio sources with calibrated flux densities provided an estimate for the efficiency of DSS 14 of 0.47. It was assumed that the other 64-m stations had the same value.

Figure 1 shows the distribution on the sky of all of the sources observed and the trajectory of the VEGA spacecraft. A correlated flux density of about 0.4 Janskys (Jy) is required for  $\Delta$ DOR observations when the antenna system temperature is 40 Kelvin as it is for the 1.7 GHz system. The sources with a correlated flux density of greater than 0.4 Jy are plotted in Fig. 1 with the region within 10 deg of the source marked.

Gaps occur in the coverage along the spacecraft trajectory for several reasons. First, there is a lack of known sources in the direction of the galactic center and anti-center which

causes the gaps around 19 and 6 hours Right Ascension, respectively. A program is currently underway (Ref. 6) to identify suitable navigation sources in these regions. Secondly, observations could not be scheduled for the region from 13 to 16 hours R.A. due to conflicts with other projects. Finally, the sun, which moves on a path very close to the spacecraft path as seen from the Earth, prevented the observation of several sources that would otherwise have been included in the list.

Table 1 lists all of the observed sources along with their correlated flux densities as determined by these measurements. The range of interferometric hour angle for the observations of each source is also listed. Some differences in flux densities were observed between the two baselines and so both are listed in Table 1. Very little variation in flux density with IHA was observed over the limited ranges of IHA observed and thus only the average value is presented in Table 1.

The accuracy of the flux density measurements depends primarily on the accuracy of the system temperature measurements and the measured efficiencies of the antennas. Formal

errors in the correlated VLBI amplitudes were generally less than 5%. The error in the efficiency determination was estimated from the uncertainty in the flux density calibration of the radio sources used for calibration and from scatter in repeated measurements. The measured value was 0.47 with an uncertainty of 0.05.

The system noise temperatures (SNTs) were read off of calibrated strip charts. These charts can generally be read with a 1-deg accuracy which corresponds to a 2.5% error. However, drift in temperatures with time made the calibrations of the charts questionable and it was estimated the SNT measurements were no better than 10%.

It is therefore estimated that the statistical error on the correlated flux density measurements is 9% with a systematic error of 11% from the efficiency calibration.

Table 2 lists the positions of all of the sources observed. The positions are given in the J2000 coordinate system and come from the most recent source catalogue, 1986D2.CAT (O. J. Sovers, private communication, 1986).

## References

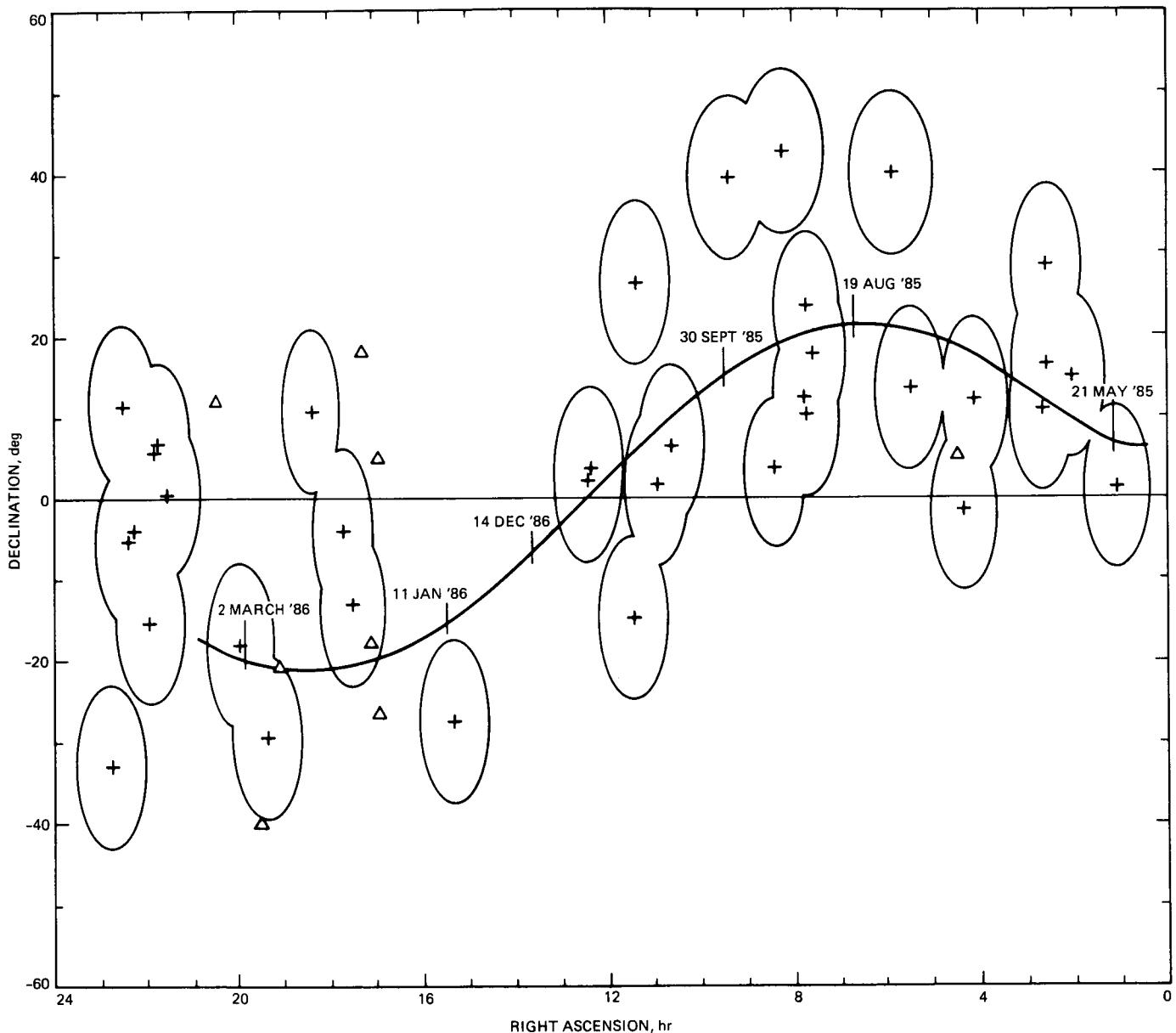
1. Thomas, J. B., "An Analysis of Long Baseline Radio Interferometry," *DSN Technical Report 32-1526*, Vol. VII, Jet Propulsion Laboratory, Pasadena, Calif., pp. 37-50, Feb. 15, 1972.
2. Thomas, J. B., "An Analysis of Long Baseline Radio Interferometry, Part II," *DSN Technical Report 32-1526*, Vol. VIII, Jet Propulsion Laboratory, Pasadena, Calif., pp. 29-38, April 15, 1972.
3. Thomas, J. B., "An Analysis of Long Baseline Radio Interferometry, Part III," *DSN Technical Report 32-1526*, Vol. XVI, Jet Propulsion Laboratory, Pasadena, Calif., pp. 47-64, June 15, 1973.
4. Melbourne, W. G., and Currkendall, D. W., "Radio Metric Direction Finding: A New Approach to Deep Space Navigation," paper presented at the AAS/AIAA Astrodynamics Specialist Conference, Jackson Hole, Wyoming, Sept. 7-9, 1977; copies are available from the author of this article, K. M. Liewer.
5. Brunn, D. L., Preston, R. A., Wu, C. S., Siegel, H. L., Brown, D. S., Christensen, C. S., and Hilt, D. E., "VLBI Spacecraft Tracking Demonstration: Part 1, Design and Planning," *TDA Progress Report 42-45*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 111-132, Mar.-Apr. 1978.
6. Ulvestad, J. S., and Linfield, R. P., "The Search for Reference Sources for delta-VLBI Navigation of the Galileo Spacecraft," *TDA Progress Report 42-84*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 152-163, Oct.-Dec. 1985.

**Table 1. Correlated flux densities at 1.7 GHz**

Source	California-Australia		California-Spain	
	I.H.A. Range, hr, min	Correlated Flux Density, Jy	I.H.A. Range, hr, min	Correlated Flux Density, Jy
P 0106+01	1 04 to 1 52	2.9	-13 08 to -12 44	3.7
P 0202+14	-0 12 to 0 36	1.2	-14 12 to -12 16	1.3
CTD 20	-0 14 (1 obs.)	1.6	-14 28 to -13 04	1.6
GC 0235+16	-0 40 to 0 12	0.50	-14 20 to -12 56	0.58
OD 166	-0 36 to 0 12	1.1	-14 04 to -12 48	1.1
GC 0406+12	-0 56 to 1 36	1.1	-14 04 (1 obs.)	0.79
3C 120	-1 16 (1 obs.)	0.25		
P 0420-01	0 48 to 1 28	2.5		
P 0528+134	-0 12 to 2 08	0.61	-10 04 (1 obs.)	0.82
DA 193	0 52 to 1 52	2.1	-10 16 (1 obs.)	2.0
P 0735+17	-0 32 to 0 56	0.58	-11 48 to -9 56	0.67
DW 0742+10	-1 36 to 0 28	1.1	-11 44 to -10 00	2.5
B2 0745+24	0 08 (1 obs.)	0.94	-12 04 to -10 12	0.42
P 0748+126	-0 40 to 0 20	0.58	-11 52 to -11 04	0.37
OJ 425			-11 52 (1 obs.)	1.1
P 0823+033	-1 56 to 0 04	0.51	-12 08 to -11 16	0.90
4C 39.25			-11 28 (1 obs.)	1.5
OL 064.5	-1 56 (1 obs.)	0.79	-12 24 to -11 36	0.54
P 1055+01	-2 04 (1 obs.)	1.1	-12 36 to -12 00	0.70
P 1123+26			-13 20 to -13 00	0.54
P 1127-14			-13 08 (1 obs.)	0.70
P 1222+037			-13 40 to -13 08	1.0
3C 273			-13 40 to -13 08	6.2
P 1519-273	0 40 (1 obs.)	0.63		
DW 1656+05	0 16 (1 obs.)	0.35	-10 48 to -10 12	0.36
P 1657-261	-0 52 (1 obs.)	0.15		
OT-111	-0 56 to 1 20	0.17		
GC 1717+17	0 04 (1 obs.)	0.21	-11 00 (1 obs.)	0.50
NRAO 530	-1 12 to 1 00	1.9	-11 32 (1 obs.)	3.2
P 1741-038	-1 16 to 0 36	1.1	-11 48 to -10 36	1.2
P 1810+10	-0 52 (1 obs.)	0.99	-10 00 to -10 32	0.88
OV-213	-2 16 to -0 -31	<0.06	-12 12 to -11 44	0.92
OV-236	-2 44 to -1 32	3.6		
P 1933-400	-2 48 (1 obs.)	0.22		
OV-198	-2 56 to -0 56	1.2	-12 40 to -11 40	1.1
P 2029+121	-1 36 (1 obs.)	0.37	-14 00 to -12 00	0.59
P 2134+004			-13 28 to -12 44	2.5
P 2145+06			-13 48 to -13 12	0.78
OX 082			-13 40 to -13 08	0.57
OX-192	-2 44 to -2 28	2.0	-12 44 to -12 16	2.4
P 2216-03			-13 16 to -12 52	1.5
3C 446			-13 16 to -13 04	0.87
CTA 102			-13 36 to -13 00	1.0
P 2245-328	-3 28 (1 obs.)	1.3		

**Table 2. Positions of observed sources in J2000 coordinates**

Source Name	Right Ascension, hr min s			Declination, deg min s			R. A. Error, s	Dec. Error, s
P 0106+01	1	8	38.771114	1	35	0.31866	0.000033	0.00082
P 0202+14	2	4	50.413987	15	14	11.04409	0.000042	0.00088
CTD 20	2	37	52.405765	28	48	8.99121	0.000041	0.00061
GC 0235+16	2	38	38.930211	16	36	59.27431	0.000039	0.00070
OD 166	2	42	29.171033	11	1	0.72725	0.000045	0.00103
GC 0406+12	4	9	22.008811	12	17	39.84798	0.000041	0.00110
P 0420-01	4	23	15.800859	-1	20	33.06536	0.000033	0.00087
3C 120	4	33	11.095640	5	21	15.62095	0.000060	0.00146
P 0528+134	5	30	56.416886	13	31	55.15045	0.000035	0.00075
DA 193	5	55	30.805658	39	48	49.16675	0.000046	0.00055
P 0735+17	7	38	7.393860	17	42	18.99844	0.000037	0.00079
DW 0742+10	7	45	33.059626	10	11	12.69075	0.000032	0.00074
B2 0745+24	7	48	36.109378	24	0	24.10966	0.000046	0.00102
P 0748+126	7	50	52.045858	12	31	4.82675	0.000038	0.00082
OJ 425	8	18	15.999761	42	22	45.41382	0.000049	0.00061
P 0823+033	8	25	50.338476	3	9	24.51846	0.000032	0.00089
4C 39.25	9	27	3.013942	39	2	20.85004	0.000045	0.00060
OL 064.5	10	41	17.162545	6	10	16.92082	0.000035	0.00139
P 1055+01	10	58	29.605239	1	33	58.82181	0.000027	0.00081
P 1123+26	11	25	53.712000	26	10	19.97584	0.000037	0.00080
P 1127-14	11	30	7.052407	-14	49	27.38797	0.000132	0.00201
P 1222+037	12	24	52.421898	3	30	50.29074	0.000038	0.00111
3C 273	12	29	6.699702	2	3	8.59702	0.000001	0.00072
P 1519-273	15	22	37.675997	-27	30	10.78680	0.000113	0.00166
DW 1656+05	16	58	33.447795	5	15	16.43471	0.000408	0.00632
P 1657-261	17	0	53.154298	-26	10	51.72892	0.000061	0.00135
OT-111	17	9	34.345451	-17	28	53.36707	0.000061	0.00139
GC 1717+17	17	19	13.048552	17	45	6.43354	0.000186	0.00329
NRAO 530	17	33	2.705841	-13	4	49.54935	0.000034	0.00101
P 1741-038	17	43	58.856161	-3	50	4.61898	0.000030	0.00087
P 1821+10	18	24	2.855358	10	44	23.77107	0.000078	0.00378
OV-213	19	11	9.652812	-20	6	55.10818	0.000176	0.00252
OV-236	19	24	51.056123	-29	14	30.12252	0.000108	0.00149
P 1933-400	19	37	16.217392	-39	58	1.55016	0.000561	0.00543
OV-198	20	0	57.090601	-17	48	57.67473	0.000060	0.00124
P 2029+121	20	31	54.994196	12	19	41.34210	0.000127	0.00251
P 2134+004	21	36	38.586440	0	41	54.21206	0.000042	0.00097
P 2145+06	21	48	5.458714	6	57	38.60275	0.000029	0.00068
OX 082	21	51	37.875386	5	52	12.95535	0.000052	0.00124
OX-192	21	58	6.281910	-15	1	9.32776	0.000150	0.00227
P 2216-03	22	18	52.037816	-3	35	36.88158	0.000035	0.00095
3C 446	22	25	47.259361	-4	57	1.39173	0.000034	0.00091
CTA 102	22	32	36.408899	11	43	50.90455	0.000035	0.00076
P 2245-328	22	48	38.685744	-32	35	52.18476	0.000147	0.00181



**Fig. 1. Distribution of radio sources observed as potential 1.7 GHz VLBI reference sources. Sources with sufficient flux density for use in ΔDOR navigation are plotted as a cross. Those regions within a 10 degree radius of these sources are enclosed by a solid line. Open triangles mark those sources with insufficient flux density. The spacecraft trajectory is shown for the period of May 1985 to March 1986.**

## ICE Second Halley Radial: TDA Mission Support and DSN Operations

N. A. Fanelli

TDA Mission Support and DSN Operations

L. Efron and R. J. Muellerschoen

Navigation Systems Section

*This article documents the operations encompassing the International Cometary Explorer (ICE) second Halley radial experiment centered around March 28, 1986. The support was provided by the Deep Space Network (DSN) 64-meter subnetwork. Near continuous support was provided the last two weeks of March and the first two weeks of April to insure the collection of adequate background data for the Halley radial experiment. During the last week of March, plasma wave measurements indicate that ICE was within the Halley heavy ion pick-up region.*

### I. Introduction

The International Sun Earth Explorer (ISEE-3) was launched in 1978. Five years were spent in Earth orbit, measuring particles and fields; first monitoring the solar wind upstream of the Earth and later in the geomagnetic tail region. During this period, support of ISEE-3 was provided by the 26-m antennas operated by the Ground Spaceflight Tracking and Data Network (GSTDN). In December 1983, using a 119-km-altitude, lunar-gravity-assist swingby, the spacecraft was sent on a new mission to intercept Comet Giacobini-Zinner (G-Z). At perilune, the spacecraft was renamed the International Cometary Explorer (ICE).

Mission Trajectory acrobatics and navigation from launch through comet encounter are covered in detail in Reference 1. The Comet G-Z Encounter operations are covered in References 2-7.

While in Earth orbit, the geocentric range to ISEE-3 had been on the order of 0.01 AU ( $1.5 \times 10^6$  km) or less. While enroute to encounter with G-Z the geocentric range to ICE grew to 0.51 AU. In January 1984, ICE went out of range of the 26-m antennas and required support from the DSN 64-m antennas for the remainder of the mission.

The primary mission goal was to make the first *in situ* measurement of a comet. This was achieved on September 11, 1985 when the spacecraft passed unscathed through the G-Z tail 7800 km downstream of the nucleus.

A secondary mission goal was to obtain measurements in the solar wind upstream of Halley's Comet for possible correlation with Earth- and space-based observations of that celestial visitor during the same time frame. Twice during interplanetary cruise the ICE orbit placed the spacecraft in near

radial alignment between the sun and Halley's Comet. This occurred first in October–November 1985 and again late in March 1986. This is illustrated in Fig. 1. The second Halley radial alignment occurred when ICE approached within  $30 \times 10^6$  km of the comet. Around this time, the plasma wave instrument aboard ICE made measurements indicative of ions produced by heavy neutrals emitted by Halley's Comet. Hence, ICE can lay claim to being the sixth member of the Halley Armada, behind Sakigake, Suisei, Giotto, Vega 1 and Vega 2.

## II. Spacecraft Systems

The ICE spacecraft is a 16-sided cylinder which is 1.61 meters tall and 1.74 meters wide. Spacecraft mass at launch was 479 kg which included 104 kg of instruments and 89 kg of hydrazine propellant. There are two solar arrays, above and below an equipment platform where most of its payload of scientific instruments is mounted (Fig. 2). The solar array provides for all of the power requirements of the spacecraft.

The propulsion system is arranged in two independent and redundant systems consisting of 12 jet thrusters and 8 fuel tanks. This system is used for all spin-rate-change trajectory correction, and attitude control maneuvers. The spacecraft rotates around the central axis of the cylinder at a rate of approximately 20 revolutions per minute.

The spacecraft carries 13 scientific instruments. Several of these utilize appendages to measure conditions near the spacecraft without spacecraft interference. Magnetic fields are measured by two search coils located at the end of two 3-meter-long booms that extend in opposite directions out from the cylinder. Also, long wires, acting as radio mapping antennas, extend 92 meters tip-to-tip in the radial direction and 14 meters in the axial direction. The experiments aboard the spacecraft have directly affected the strategy for targeting the spacecraft for its Comet G-Z intercept and Halley radial experiment.

## III. Operations

The DSN utilized the 64-meter subnetwork with stations located at Goldstone, California; Madrid, Spain; and Canberra, Australia. The 64-meter subnetwork provided nearly continuous support from March 15, 1986 to April 15, 1986. The subnetwork was used to collect spacecraft telemetry data, send commands to spacecraft, and generate radiometric data. Two-way range and range rate data over a 6-week time span were utilized by the JPL Multimission Navigation Team in determining the spacecraft orbit relative to Halley's Comet. The data arc covered the period between spacecraft trajectory

correction maneuvers performed on February 27, 1986 and April 7, 1986.

High quality *in situ* Halley Comet data was obtained via the 64-meter subnetwork and sent to the ICE Project located at the Goddard Space Flight Center (GSFC) where the data was processed and provided to the science investigators. The ICE Project was very pleased with the high quality data that was provided by the DSN and as a direct result were able to obtain excellent day-to-day data correlation for the Halley radial encounter period.

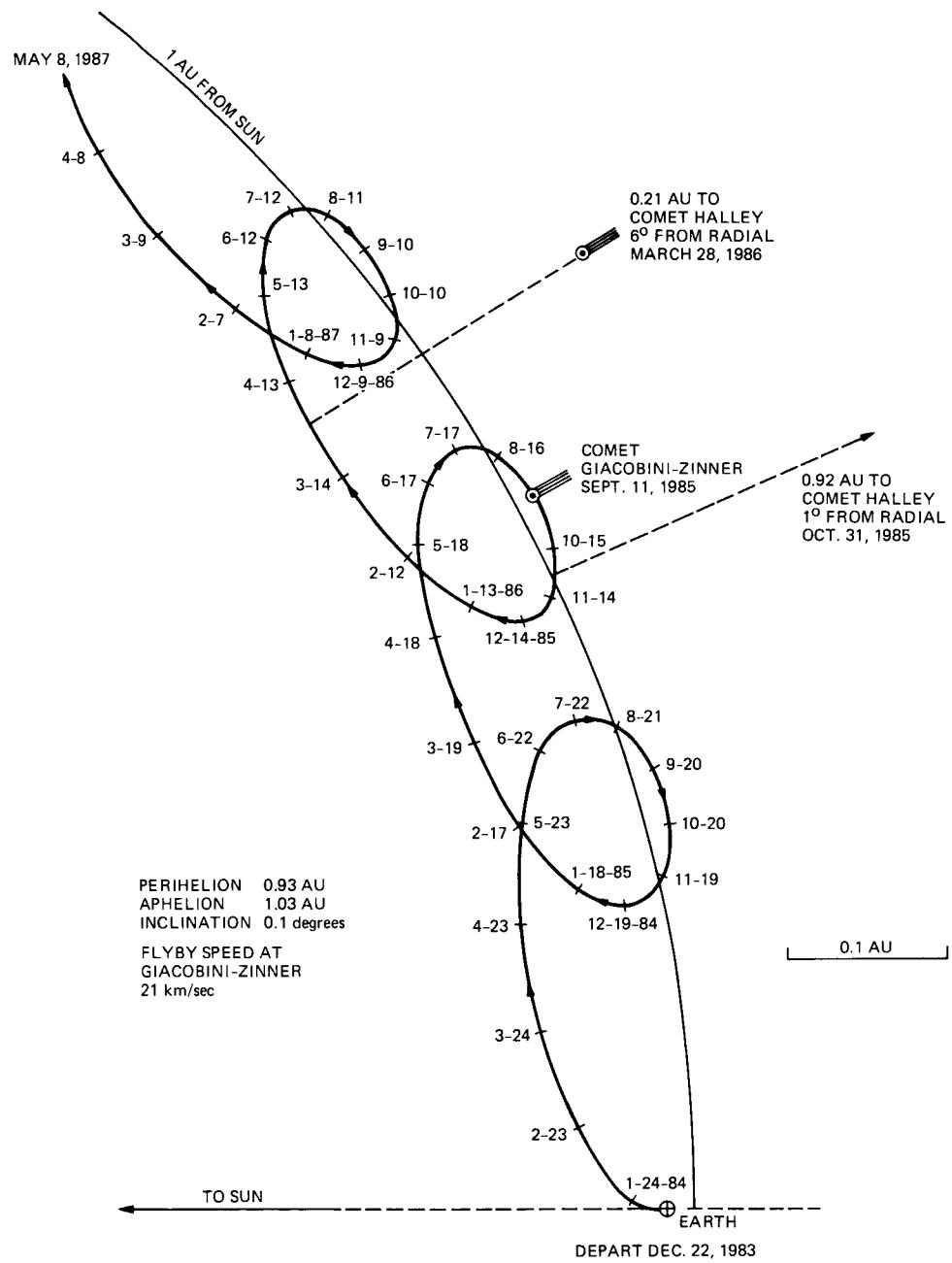
## IV. Surprising Results

Solar plasma flowing radially outward from a rotating sun, carries along with it the interplanetary magnetic field (IMF). This results in a pinwheel-like structure to the solar wind. For a plasma flowing radially outward at 450 km/sec, Fig. 3 schematically illustrates the heliocentric orbits of the Earth, ICE and Halley relative to IMF field lines at  $45^\circ$  intervals during the six month period from 1 January 1986 through 1 July 1986. Figure 4 provides three orthogonal projections of the ICE orbit in a Halley-centered reference frame. Figure 4a shows a projection for an observer above the Halley orbit plane, 4b for an observer looking toward the Sun along the Halley-Sun line, and 4c for an observer in the Halley orbit plane looking toward the Sun-Halley line. In each view the circle is the projection of a sphere of radius  $35 \times 10^6$  km centered at Halley. Due to the negligible gravity potential and weak bow shock associated with comets, neutral heavy particles emitted from the nucleus are able to travel great distances in all directions before they are ionized by the Sun's ultraviolet radiation. Once ionized, they are picked up by the plasma and begin to gyrate around the field lines.

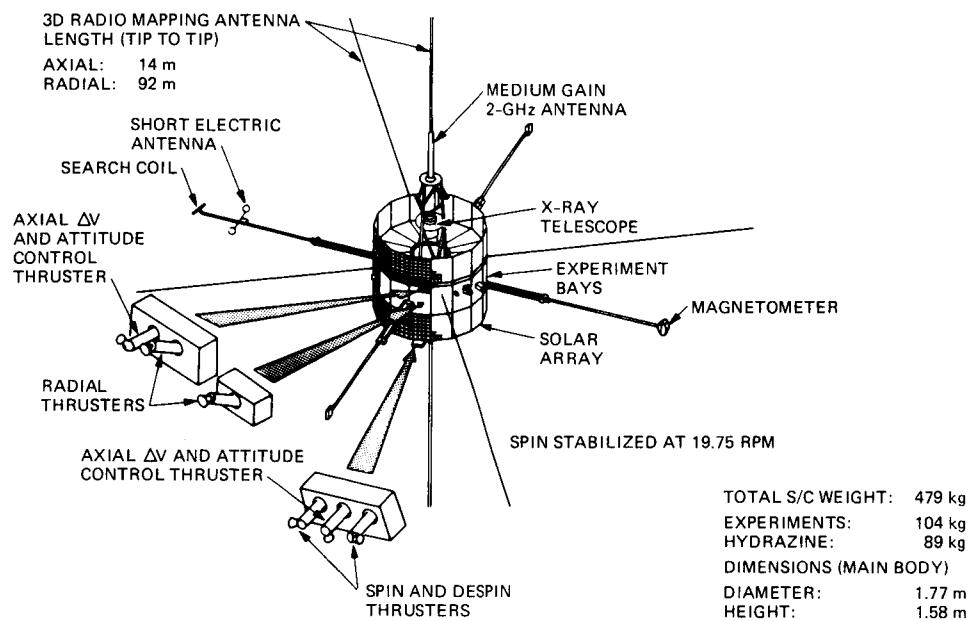
Closest approach of ICE to Halley was at a distance of  $28 \times 10^6$  km on 25 March 1986. The minimum Halley-Sun-ICE angle of about  $6^\circ$  occurred on 28 March 1986. From References 7 and 8, the energetic ion instrument (EPAS) measurements aboard ICE appear to be consistent with the interaction of the solar wind with heavy ions speculated to have their source in Halley's nucleus. Such an interpretation is in agreement with the ICE observations during the Comet G-Z encounter period. This implies a Halley's Comet solar wind interaction region extending perhaps  $30\text{--}40 \times 10^6$  km from the comet's nucleus. Hence, when Halley is 1 AU from the Sun, its plasma interaction region extends from near the orbit of Venus out to mid-way between the orbits of Earth and Mars. The ICE mission has unexpectedly revealed that comets, despite their small nuclei, can have interaction zones near perihelion which exceed in size and volume the similar zones associated with the heliocentric motion of even the giant outer planets.

## References

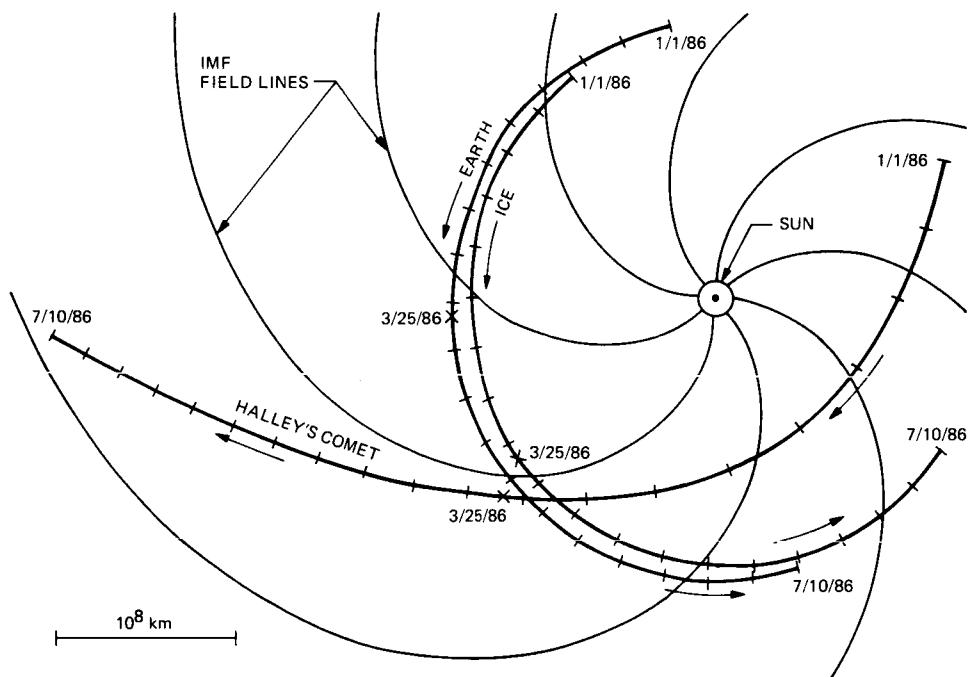
1. Efron, L., Muellerschoen, R. J., and Premkumar, R. I., "ICE Navigation Support," *TDA Progress Report 42-86*, pp. 268-283, Jet Propulsion Laboratory, Pasadena, Calif., August 15, 1986.
2. Fanelli, N. A., and Morris, D. G., "ICE Encounter Operations," *TDA Progress Report 42-84*, pp. 176-185, Jet Propulsion Laboratory, Pasadena, Calif., February 15, 1986.
3. Goodwin, J. P., "Usuda Deep Space Center Support to ICE," *TDA Progress Report 42-84*, pp. 186-196, Jet Propulsion Laboratory, Pasadena, Calif., February 15, 1986.
4. Gordon, D. D., and Ward, M. T., "Arecibo Observatory Support of the International Cometary Explorer Mission Encounter at Comet Giacobini-Zinner," *TDA Progress Report 42-84*, pp. 197-202, Jet Propulsion Laboratory, Pasadena, Calif., February 15, 1986.
5. Layland, J. W., "ICE Telemetry Performance," *TDA Progress Report 42-84*, pp. 203-213, Jet Propulsion Laboratory, Pasadena, Calif., February 15, 1986.
6. Hurd, W. J., Pollara, F., Russell, M. D., Siev, B., and Winter, P. U., "Intercontinental Antenna Arraying by Symbol Stream Combining at Giacobini-Zinner Encounter," *TDA Progress Report 42-84*, pp. 220-228, Jet Propulsion Laboratory, Pasadena, Calif., February 15, 1986.
7. Scarf, F. L., et al., "ICE Plasma Wave Measurements in the Ion Pick-up Region of Comet Halley," *Geophysical Research Letters*, Vol. 13, No. 8, pp. 857-860, August 1986.
8. Wenzel, K. P., et al., "In-Situ Observations of Cometary Pick-Up Ions  $\geq 0.2$  AU Upstream of Comet Halley: ICE Observations," *Geophysical Research Letters*, Vol. 13, No. 8, pp. 861-864, August 1986.



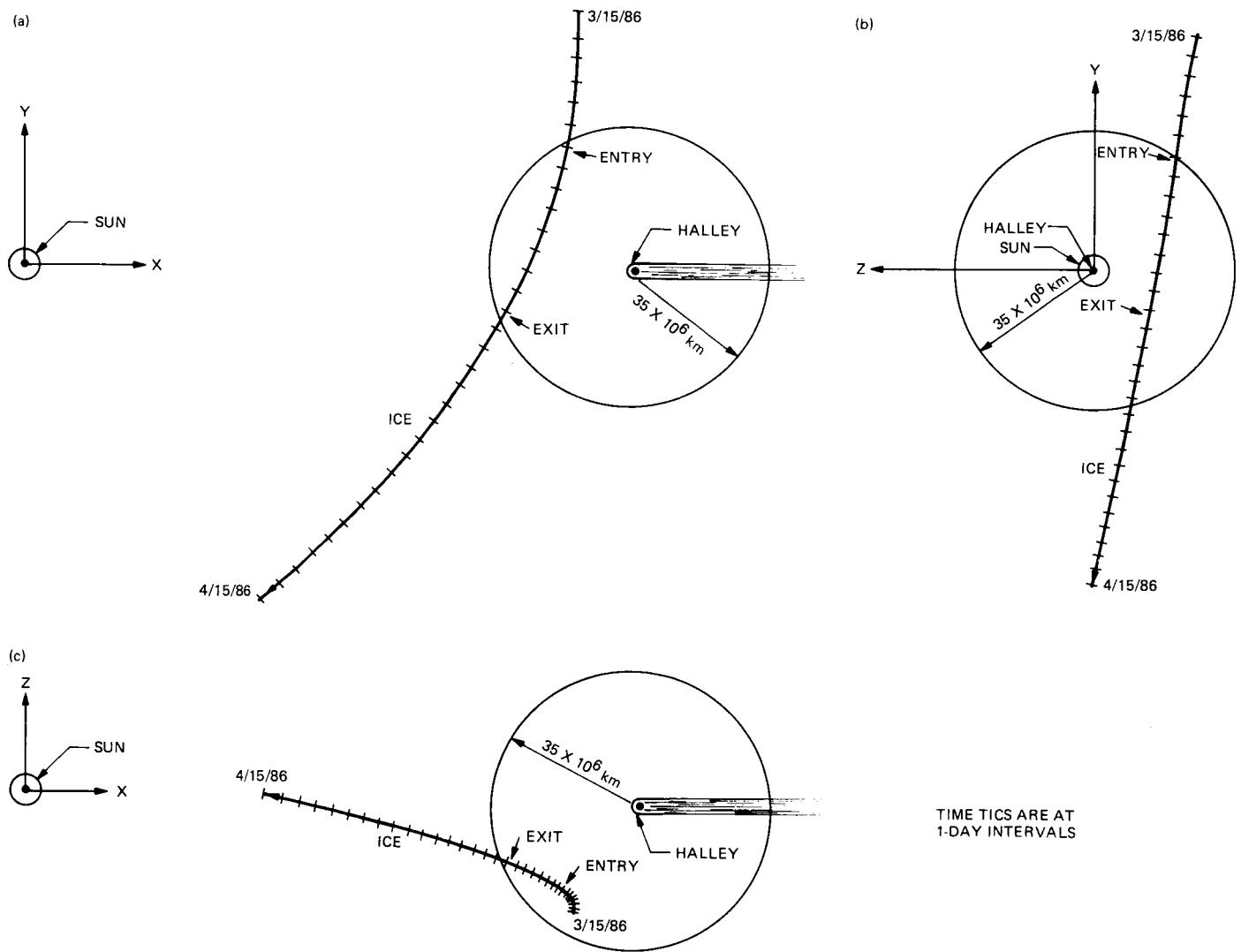
**Fig. 1. ICE trajectory relative to a fixed Sun-Earth line**



**Fig. 2. The ICE spacecraft**



**Fig. 3. Heliocentric orbits of ICE, Halley's Comet and Earth projected onto the ecliptic plane, 1 Jan. 1986 to 10 July 1986. Time ticks are at 10-day intervals. IMF field lines illustrated schematically.**



**Fig. 4.** Halley-centered orbit of ICE (15 March 1986 to 15 April 1986) and the projection of a  $35 \times 10^6$  km radius sphere about the nucleus, as viewed by an observer (a) above the comet orbit plane, (b) looking toward the Sun along the Halley-Sun line from within the comet tail, and (c) in the comet orbit plane looking normal to the Sun-Halley line

D30-32  
P9

July—September 1986

N87-15359

## Propagation Through Martian Dust at 8.5 and 32 GHz

E. K. Smith

Telecommunications Systems Section

W. L. Flock

University of Colorado

*Independent studies of attenuation of X-band (8.5 GHz) and Ka-band (32 GHz) radio signals when traversing Martian dust were carried out by the authors. These analyses turned out remarkably similar. The computational method is essentially that of T. S. Chu but uses observed optical depth at 0.67 microns rather than "visibility" as the measure of optical attenuation from which to derive the microwave attenuation. An awkwardness in the approach is that the size distribution of Martian dust particles is not well known, but the mean is probably around 4 microns, whereas in the terrestrial case it is nearer 10 microns. As a consequence, there will be a larger tail of particles still in the Mie regime in the Martian case as compared to the terrestrial one. The computational error will, therefore, be somewhat larger for Martian than Earth-bound dust. Fortunately, the indicated attenuations are small enough for the worst case (1.3 dB at 32 GHz) that the error is academic.*

### I. Introduction

This article makes use of the approaches and results of studies of terrestrial sand and dust storms for the rather analogous problem on Mars. Dust storms on Mars are, if anything, more impressive than on Earth. Typically, in spring in the southern hemisphere, dust storms will start, grow, move northward, and frequently encircle the globe. In 1971, there was a huge storm which obliterated the surface features of the entire planet.

the luminance  $L$  of the object has decreased to 0.031 of its initial value  $L_0$ . As luminance is a measure of power, this translates to:  $L/L_0 = 10 \log 0.031 = 15.086$  dB, which is usually shortened to 15 dB. An approach developed by Chu (Ref. 6), and used by Ansari and Evans (Ref. 1) and Goldhirsh (Ref. 8), incorporates the visibility  $V$  into the expression for microwave attenuation coefficient  $A(2)$ :

$$A(\lambda) = \frac{189}{V(\text{km})} \frac{r}{d} \left[ \frac{3\epsilon_i}{(\epsilon_r + 2)^2 + \epsilon_i^2} \right] \text{dB/km}$$

where

$r$  = particle radius, m

$\epsilon = \epsilon_r - j\epsilon_i$  = complex dielectric constant

$\lambda$  = wavelength, m

As shown above, particles are assumed to be uniform in size and to have known permittivity. The development assumes single scatter. This is not bad at microwave, but the optical visibility limit is well into the multiple scatter regime.

Particle size distributions for four dust storms in Khartoum have been published by Ghobrial (Ref. 7), who finds particles with sizes ranging from 0.1 to 300 microns (0.0001 to 0.3 mm). Wind tunnel measurements (Ref. 9) indicate that 0.08 mm is the most likely size of particle to be picked up by wind from the Earth's surface.

### III. Mars

Mars has long captured the public fancy as possibly the most hospitable planet, next to Earth, for life. This is not so unreasonable, as can be seen in Table 1. However, the temperature is considerably colder, the atmospheric pressure at the surface less than 1% that of the Earth, and the atmosphere almost totally lacking in oxygen (95.3% CO<sub>2</sub>, 2.7% N<sub>2</sub>, 1.6% argon and perhaps 0.13% O<sub>2</sub>; Ref. 13). Mars' orbit around the sun is considerably more eccentric than that of the Earth (206.7 × 10<sup>6</sup> km at perihelion and 249.1 × 10<sup>6</sup> km at aphelion) with the result that the insolation is some 45% greater during perihelion (late spring in the southern hemisphere) than at aphelion (late spring in the northern hemisphere).

Mars has been visited by spacecraft from both the US and USSR as is seen in Table 2. The most recent visits were by the two US Viking missions, and the data presented here came largely from that source.

Martian dust storms tend to originate in the late spring or summer in the southern hemisphere when the Martian surface temperatures and the temperature gradients are highest. The greatest storm on record started in September of 1971 and became planetwide (Ref. 11). Great storms (planetary scale) occur once or twice each Martian year (Ref. 19). Quite typically, a storm will move northward and reach planet-encircling proportions and last several weeks. The dust has been reported as high as 50 km above the surface. Optical depths as great as 6 at 0.67 microns have been measured and a dust density scale height of 10 km proposed (Refs. 15 and 16). The size and permittivity of Martian dust remain controversial. Figure 1 portrays a particle density distribution (Ref. 10) typical of those found in papers published in the late 1970s. The mean of the particle radius is 0.4 microns. Also shown is the geometric cross-section distribution, GN(r), which is more directly pertinent to extinction calculations. Recent workers have favored somewhat larger particles, around 2 microns (Ref. 19 and Footnote 1) or greater (Ref. 5). The favored substances for the dust are basalt and montmorillonitic clays.

Tabulated dielectric constant at 50 microns (the longest wavelength available) for the latter is 2.18 - j0.14.<sup>1</sup>

Water was detected in the soils at the two Viking Lander sites, but the integrated water in the Martian atmosphere does not exceed 100 precipitable microns (Ref. 4).

### IV. Attenuation by Dust

A simple slab model of a Martian dust storm is shown in Fig. 2. The assumption is made that the reflected components of the electric field  $E_r$  and  $E_{tr}$  are much less than the transmitted component  $E_t$  ( $E_t \gg E_r, E_t \gg E_{tr}$ ). We define a complex propagation constant  $K_c$ :

$$K_c = \beta - j\alpha \quad (1)$$

where

$$\beta = K_0 n_r = \frac{2\pi n_r}{\lambda_0} = \text{phase constant} \quad (2)$$

$$\alpha = K_0 n_i = \frac{2\pi n_i}{\lambda_0} = \text{attenuation constant} \quad (3)$$

and

$$n_i = n_r - jn_i = \text{complex refractive index}$$

$$K_0 = \frac{2\pi}{\lambda_0}$$

where

$$\lambda_0 = \text{free-space wavelength}$$

One can then write for the electric field (Ref. 2)

$$E = E_t(0, 1, 0) \exp \{j(\omega t - Kz)\} \quad (4)$$

where  $E_t(0, 1, 0)$  is the complex amplitude of the transmitted plane wave with electric vector polarized in the  $y$  direction and travelling downward (positive  $z$  direction). Substituting (1) into (4) yields

$$E = E_t(0, 1, 0) \exp(-\alpha z) \exp \{j(\omega t - \beta z)\} \quad (5)$$

<sup>1</sup>R. A. West, Jet Propulsion Laboratory, private communication, 1986.

which may also be written

$$E = E_t(0,1,0) \exp\left(-\frac{1}{2}\tau\right) \exp\left\{j\left(\omega t - \frac{2\pi z}{\lambda}\right)\right\} \quad (6)$$

where  $\tau$  is the optical depth. The attenuation coefficient  $A$  in decibels is then given by:

$$A \text{ (dB)} = 8.686 \alpha \text{ (nepers)} \quad (7)$$

$$A \text{ (dB)} = 8.686 K_0 n_i \text{ dB/m} \quad (8)$$

$$\text{Attenuation from dust} = 4.343\tau \text{ dB} \quad (9)$$

### A. Chu's Development

We now turn to the work of Chu (Ref. 6) who wrote a much quoted article on "Effects of Sandstorms on Microwave Propagation." Chu considered sand of radius 0.01 to 0.1 mm (10 to 100 microns) with dielectric constant 2.5 ( $1 - j0.01$ ) to 10 ( $1 - j0.01$ ). He invoked the Rayleigh approximation at microwave and the very large sphere approximation in the visual range. The attenuation coefficient  $\alpha$  is defined as

$$\alpha = \int_0^{\infty} N(a)C(a) da \quad (10)$$

where

$C(a)$  = extinction coefficient

$N(a)$  = number density

$a$  = particle radius

For sand of radius  $a$

$$\alpha = 4.34 \left[ \frac{S}{(4/3)\pi a^3} \right] (\pi a^2 Q_{\text{ext}}) = \frac{3.25 S Q_{\text{ext}}}{a} \text{ dB/m} \quad (11)$$

where  $Q_{\text{ext}} = C/\pi a^2$  is the extinction efficiency, and  $S = (4/3)\pi a^3 N$  is the fraction of the volume actually made up of sand. As the size of particles become large with respect to wavelength,  $Q_{\text{ext}}$  becomes asymptotic to 2 (Refs. 17 and 18):

$$Q_{\text{ext}} = 2, a \gg \lambda \quad (12)$$

As can be seen in Fig. 3, if the sphere is at all conducting, the extinction becomes very close to 2 by the time  $2\pi a/\lambda = 15$  or  $a = 2.4\lambda$ . Hence, in the case treated by Chu, if we take 0.6 microns for the wavelength of white light, any particle larger than 1.5 microns would meet the requirements of (12).

The question of multiple scatter needs to be raised. Van de Hulst (Ref. 18) suggests that whenever the optical depth exceeds 0.3 multiple scattering becomes a factor. In this case, the limit of visibility occurs at  $\tau = 3.47$ , well into the multiple scatter regime. However, this limit occurs at the point where a mark disappears into the background, and it can be argued that this is the coherent component that one is tracking and that single scatter still applies.

The introduction of (12) is done by combining (11) and (12) and solving for  $N$ :

$$N = \frac{\alpha_0 a}{6.5 [(4/3)\pi a^3]} \quad (13)$$

where  $\alpha_0$  is the optical attenuation coefficient in dB/m.

The next step is to make use of an expression presented, but not derived, by Van de Hulst (Ref. 18) for the effective refractive index  $m$  of a scattering medium:

$$\bar{m} = 1 - iS(0) 2\pi N K^{-3} \quad (14)$$

where  $S(0)$  is the forward scattering function, and  $K$  is the free space phase constant. In the Rayleigh regime  $S(0)$  is given by

$$S(0) = iK^3 \left( \frac{\epsilon - 1}{\epsilon + 2} \right) a^3 + \frac{2}{3} K^6 \left( \frac{\epsilon - 1}{\epsilon + 2} \right)^2 a^6 \quad (15)$$

The second term is negligible for  $\lambda \gg a$  as is the case at radio frequencies. Substituting (13) and (15) into (14) and using the terminology of (1) through (9), the phase shift is given by

$$K(R_e \bar{m} - 1) = \frac{3K}{13} \alpha_0 a \left[ R_e \left( \frac{\epsilon - 1}{\epsilon + 2} \right) \left( \frac{180}{\pi} \right) \right] \text{deg/m} \quad (16)$$

and the attenuation coefficient becomes

$$K(\text{Im } \bar{m}) = \frac{3K}{13} \alpha_0 a \left[ \text{Im} \left( \frac{\epsilon - 1}{\epsilon + 2} \right) \right] (8.68) \text{ dB/m} \quad (17)$$

### B. Application to the Martian Problem

For a uniform slab the total attenuation and total phase shift for the wave of relations (4) through (6) is obtained by multiplying (16) and (17) by  $z$ , the distance traversed. For total attenuation (17) becomes

$$K(\text{Im } \bar{m})z = \frac{3K}{13} (\alpha_0 z) a \left[ \text{Im} \left( \frac{\epsilon - 1}{\epsilon + 2} \right) \right] (8.68) \text{ dB/m} \quad (18a)$$

and, substituting  $4.34\tau$  for  $\alpha_0 z$ , we may write

$$K(\text{Im } \bar{m}) = \frac{3K}{13} \tau a \left[ \text{Im} \left( \frac{\epsilon - 1}{\epsilon + 2} \right) \right] (37.7) \text{ dB/m} \quad (18b)$$

This relation (18b) now gives the microwave attenuation for uniform spheres of known radius  $a$  and dielectric constant  $\epsilon$  for a slab of constant particle density  $N$  and agrees with Eq. (13) of Goldhirsh (Ref. 8). Martian dust undoubtedly follows some particle distribution such as shown in Fig. 1, but the uncertainty as to the mean size is such that the assumption of a single particle radius is as good as any. Similarly, the dielectric constant can only be guessed at, but can reasonably be bounded. The dust density probably follows an experimental decay with height, so the equivalent slab will be used.

### C. Equivalent Slab for a Martian Dust Storm

Pollack (Ref. 15) has suggested a scale height for dust particle density,  $H_d$ , of 10 km. Assuming  $H_d$  is a constant over the region of interest, we may write

$$\int_0^\infty N_0 \exp \left( -\frac{h}{H_d} \right) dh = N_0 H_d \quad (19)$$

where the right-hand side is obtained by performing the indicated integration; and  $N_0$  is the surface density of the dust. Thus, it is seen that the Martian dust storm may be represented by a slab of height  $H_d = 10$  km and density  $N_0$ . A spherical model of such a slab is shown in Fig. 4.

### D. Calculation of Microwave Attenuation

Performing the indicated complex arithmetic on (18b) leads to

$$A = 54.62 \frac{\tau a}{\lambda} \left[ \frac{3\epsilon_i}{(\epsilon_r + 2)^2 + \epsilon_i^2} \right] \text{ dB} \quad (20)$$

where  $\epsilon = \epsilon_r - j\epsilon_i$  and  $a$ , the particle radius, and  $\lambda$  are in meters.

For zenith angles of  $0^\circ$  to  $80^\circ$  we use:

$$\tau = \tau_z \sec \theta, \quad \theta < 80^\circ \quad (21)$$

For zenith angles greater than  $80^\circ$  we need to consider refraction effects. The Martian atmosphere is 95%  $\text{CO}_2$ ;  $\text{CO}_2$  at  $0^\circ\text{C}$  and 1 atmosphere and for white light has an index of refraction of 1.000450. In terms of refractivity  $N = (n - 1)10^6$ ; this may be written

$$N(\text{CO}_2) = 121.3 p(\text{mb})/T(\text{K}) \quad (22)$$

If  $p = 6$  mb and  $T = 280$  K on the Martian surface, the surface refractivity is  $2.6N$  units as compared to about 300 on the surface of Earth, and refractive bending will be a second-order effect and can be ignored. Using the geometry in Fig. 4 for zenith angles  $\theta$  greater than  $80^\circ$  or elevation angles  $\phi$  between  $0^\circ$  and  $10^\circ$ , the optical depth  $\tau$  is

$$\tau = \frac{\tau_z}{H_d} \left\{ [(H_d + r_0)^2 - (r_0 \cos \phi)^2]^{1/2} - r_0 \sin \phi \right\},$$

$$\text{for } \phi = 90^\circ - \theta < 10^\circ \quad (23)$$

Shown in Fig. 5 is the zenith angle dependence for the "best guess" and "worst case" calculations using relations (20) and (23). Dust storms under the worst case scenario only become significant for ray paths nearly tangential to the surface of Mars. Shown in Fig. 5 is the dependence of attenuation on choice of particle radius and dielectric constant.

### V. Discussion

Attenuation through Martian dust storms has been computed for 8.5 and 32 GHz one-way radio transmission for the maximum expected optical depth ( $\tau = 6$ ). These are first-order calculations based on single scatter. Due to uncertainties in the Martian dust parameters – particle size distribution and dielectric constant – "best guess" and "worst case" situations have been selected. Attenuation, even at 32 GHz, will not be significant, except for rays almost tangential to the Martian surface.

## References

- Ansari, A. J., and Evans, B. G., "Microwave Propagation in Sand and Dust Storms," *Proc IEE*, Vol. 129, Part F, pp. 315–322, Oct. 1982.

2. Booker, H. G., "Energy in Electromagnetism," *IEE Electromagnetic Waves*, Series 13, Peter Peregrinus, Ltd., 1982.
3. Carr, M. H., *The Surface of Mars*, Yale University Press, 1981.
4. Carr, M. H., *The Geology of the Terrestrial Planets*, NASA SP-469, Supt. of Documents, 1984.
5. Christensen, P. R., "Regional Dust Deposits on Mars: Physical Properties, Age and History," *J. Geophys. Res.*, Vol. 91, No. B3, pp. 3533-3545, 1986.
6. Chu, T. S., "Effects of Sandstorms on Microwave Propagation," *Bell Syst. Tech. J.*, Vol. 58, pp. 549-555, Feb. 1979.
7. Ghobrial, S. F., "The Effect of Sandstorms on Microwave Propagation," *Proc. Nat. Telecomm. Conf.*, Houston, TX, Vol. 2, pp. 43.5.1-43.5.4, 1980.
8. Goldhirsh, J., "A Parameter Review and Assessment of Attenuation and Backscatter Properties Associated with Dust Storms Over Desert Regions in the Frequency Range of 1 to 10 GHz," *IEEE Trans. Ant. Propag.*, Vol. AP-30, pp. 1121-1127, Nov. 1982.
9. Greeley, R., Leach, R., White, R., Iverson, J., and Pollack, J., "Threshold Wind Speeds for Sands on Mars: Wind Tunnel Simulations," *Geophys. Res. Lett.*, Vol. 7, pp. 121-124, 1980.
10. Hunt, G. E., "On the Opacity of Martian Dust Storms Derived by Viking IRTM Spectral Measurements," *J. Geophys. Res.*, Vol. 84, pp. 8301-8310, 1979.
11. Martin, L. J., "Clearing the Martian Air: The Troubled History of Dust Storms," *Icarus*, Vol. 57, pp. 317-321, 1984.
12. Middleton, W. E. K., *Vision Through the Atmosphere*, University of Toronto Press, Canada, 1952.
13. Owen, T., Biemann, K., Rushneck, D. R., Biller, J. E., Howarth, D. W., and Lafleur, A. L., "The Composition of the Atmosphere at the Surface of Mars," *J. Geophys. Res.*, Vol. 82, pp. 4635-4639, 1977.
14. Pollack, J. B., *The Solar System*, Chap. 8, Scientific American, 1975.
15. Pollack, J. B., Colburn, D., Kahn, R., Hunter, J., Van Camp, W., Carlston, C. E., and Wolf, M. R., "Properties of Aerosols in the Martian Atmosphere, as Inferred from Viking Lander Imaging Data," *J. Geophys. Res.*, Vol. 82, pp. 4479-4496, 1977.
16. Pollack, J. B., Colburn, D., Flaser, M., Kahn, R., Carlston, C. E., and Pidek, D., "Properties and Effects of Dust Particles Suspended in the Martian Atmosphere," *J. Geophys. Res.*, Vol. 84, pp. 2929-2945, 1979.
17. Ulaby, F. T., Moore, R. K., and Fung, A. K., *Microwave Remote Sensing: Active and Passive*, Volume I, Fig. 5.11, p. 292, Addison-Wesley Publishing Co., Reading, MA, 1981.
18. Van de Hulst, H. C., *Light Scattering by Small Particles*, John Wiley and Sons, New York, Chapman and Hall, Ltd., London, 1957.
19. Zurek, R. W., "Martian Great Dust Storms: An Update," *Icarus*, Vol. 30, pp. 288-310, 1982.

## Bibliography

- Bagnold, R. A., *The Physics of Blown Sand and Desert Dunes*, Methuen & Co., Ltd., London, 1941.
- Flock, W. L., *Telecommunications in Planetary Environments*, JPL Publication 81-84, Jet Propulsion Laboratory, Pasadena, CA, 1981.
- Flock, W. L., *Propagation Effects on Satellite Systems at Frequencies Below 10 GHz*, NASA Reference Publication 1108, NTIS, 1983.
- Ishimaru, A., *Wave Propagation and Scattering in Random Media, Vol. I: Single Scattering and Transport Theory*, Academic Press, New York, 1978.

**Table 1. Some physical properties of Earth and Mars**

Property	Earth	Mars
1. Mean distance from sun, km $\times 10^6$	149.6	227.9 { 206.7 249.1
2. Period of revolution, days	365.28	687
3. Rotation period, hours	23.9344	24.6321
4. Equatorial diameter, km	12,756	6,787
5. Mass (Earth = 1)	1	0.108
6. Density (Water = 1)	5.5	3.9
7. Atmosphere	Nitrogen Oxygen	Carbon dioxide Nitrogen, Argon
8. Water vapor (integrated column), cm <sup>2</sup>	2.2	0.001
9. Water content in surface	0 - 30%	~1%
10. Mean atmospheric pressure, mb	1013.25	6
11. Mean surface temperature, °C	22	-23
12. Mean particle size in dust storms (diam), microns	0.5 to 100	0.1 to 10
13. Real part refractive index (clay 8.5 GHz)	[2.5 + 0.5 (% H <sub>2</sub> O)]	
14. Imaginary part refractive index (clay 8.5 GHz)	[0.02 + 0.15 (% H <sub>2</sub> O)]	
15. Real part refractive index (clay 32 GHz)	[2.5 + 0.3 (% H <sub>2</sub> O)]	
16. Imaginary part refractive index (clay 32 GHz)	[0.06 + 0.1 (% H <sub>2</sub> O)]	

Notes: 1. Data for 1 to 8 from Ref. 14.  
 2. Values for 13 to 16 derived for data for clay from Table 1, Ref. 1.

**Table 2. Spacecraft missions to Mars<sup>a</sup>**

Mission	Arrival date	Type
<b>United States</b>		
Mariner 4	July 15, 1965	Flyby
Mariner 6	July 30, 1969	Flyby
Mariner 7	August 4, 1969	Flyby
Mariner 9	November 14, 1971	Orbiter
Viking 1	June 19, 1976	Orbiter/Lander
Viking 2	August 7, 1976	Orbiter/Lander
<b>Soviet</b>		
Mars 2	November 27, 1971	Orbiter/Lander
Mars 3	December 2, 1971	Orbiter/Lander
Mars 4	March 1974	Orbiter
Mars 5	March 1974	Orbiter
Mars 6	March 1974	Lander
Mars 7	March 1974	Lander

<sup>a</sup>From Ref. 3.

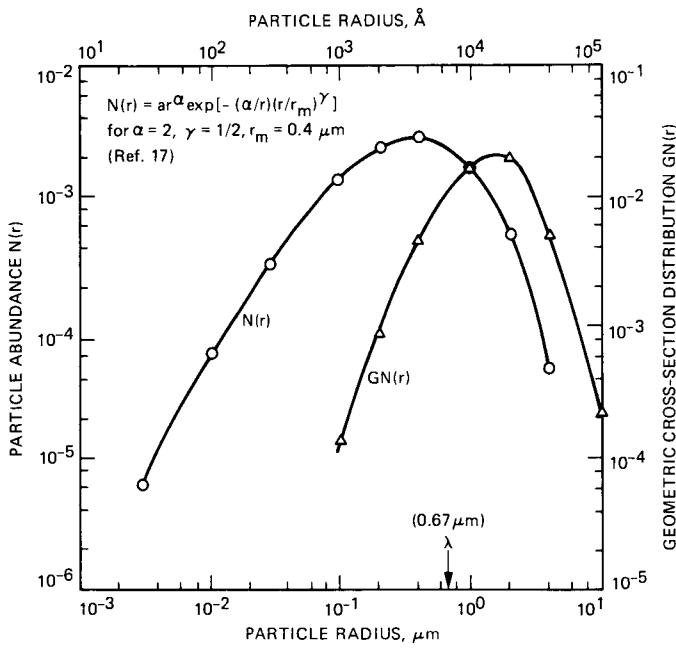


Fig. 1. Martian dust particle distribution

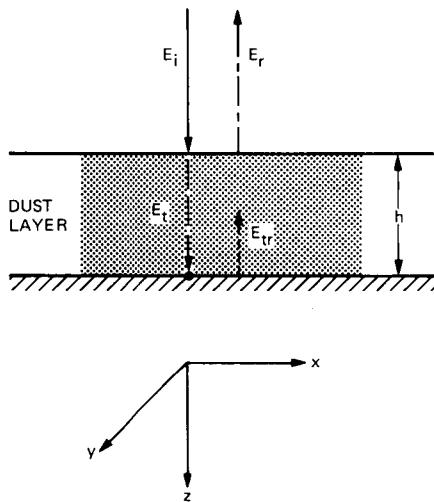


Fig. 2. Simple slab model of a Martian dust storm

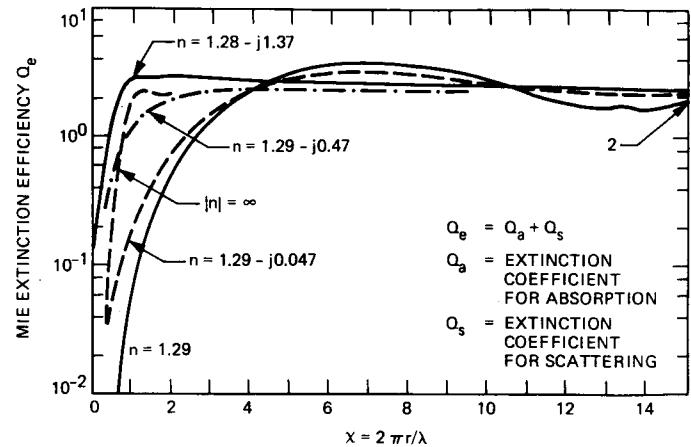


Fig. 3. The extinction efficiency  $Q_e$  for various types of spheres, shown in the range  $0 < x < 5$  (Ref. 7)

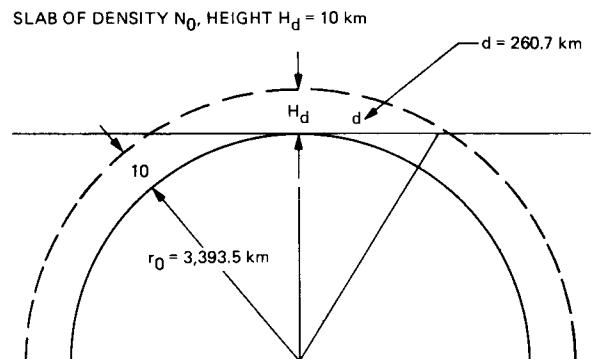


Fig. 4. Model of equivalent Martian atmosphere

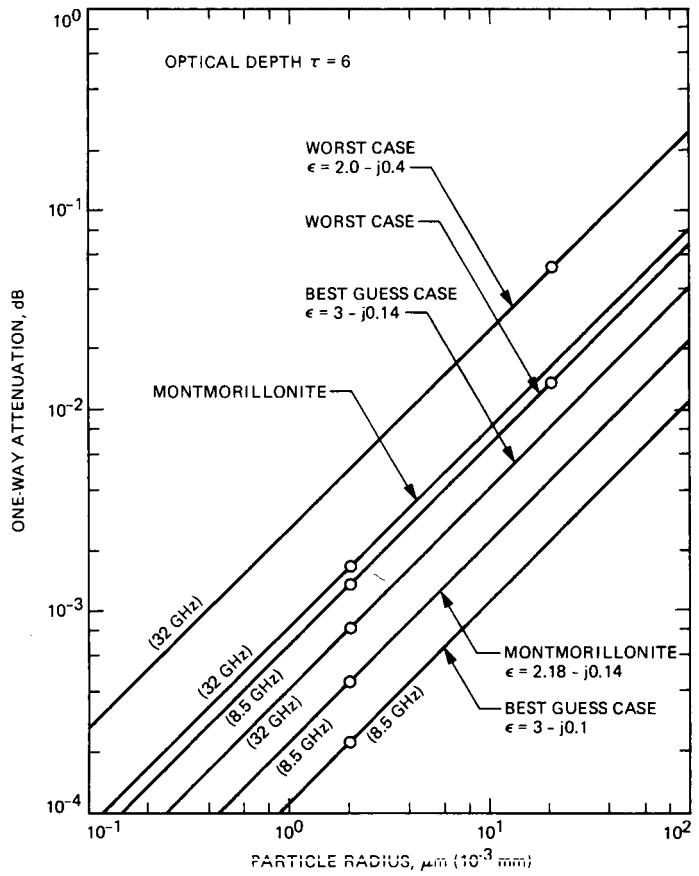
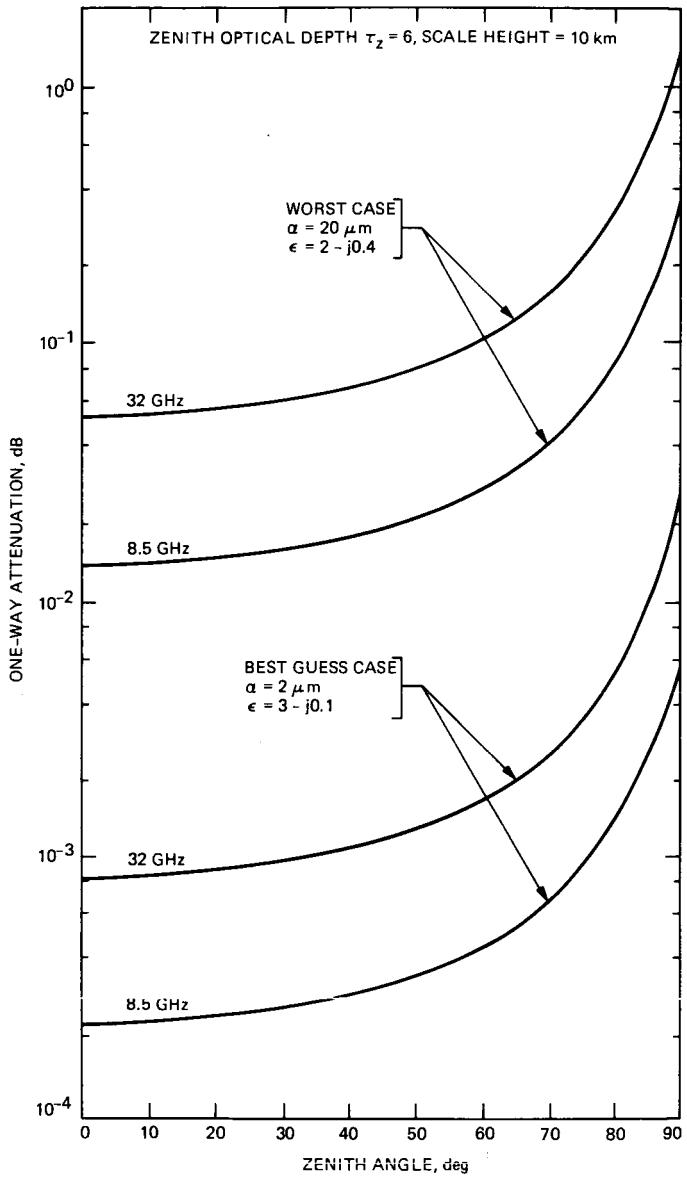


Fig. 6. Dependence of attenuation on choice of particle radius and dielectric constant

Fig. 5. Zenith angle dependence of dust attenuation in Martian

D31-33 Q M

N87-15360

TDA Progress Report 42-87

July-September 1986

## Systolic VLSI for Kalman Filters

H.-G. Yeh

California State University, Long Beach

J. J. Chang

Flight Computer Systems and Technology

*A novel two-dimensional parallel computing method for real-time Kalman filtering is presented. The mathematical formulation of a Kalman filter algorithm is rearranged to be the type of Faddeev algorithm for generalizing signal processing. The data flow mapping from the Faddeev algorithm to a two-dimensional concurrent computing structure is developed. The architecture of the resulting processor cells is regular, simple, expandable, and therefore naturally suitable for VLSI chip implementation. The computing methodology and the two-dimensional systolic arrays are useful for Kalman filter applications as well as other matrix/vector based algebraic computations.*

### I. Introduction

For more than 20 years, the Kalman filter has been applied extensively in many signal processing applications, including target tracking, adaptive controls, radar-signal processing, and failure-proof systems (Refs. 1 to 4). The applicability of the Kalman filter to real-time signal processing problems is in general limited by the relatively complex mathematical operations necessary in computing the Kalman filtering algorithm (Refs. 5 and 6). However, with the rapid development of VLSI integrated circuits (Refs. 7 to 9), it is quickly becoming technologically feasible to implement Kalman filters in real time.

The processing of a Kalman filter requires matrix/vector operations such as multiplication, addition, subtraction, and inversion. Among these, matrix inversion is the most difficult to implement in terms of speed and accuracy. The Faddeev algorithm (Refs. 10 and 11) has been suggested as a universal

algorithm for various matrix manipulations due to the fact that it is easily systematized for matrix calculations and maps easily into a concurrent systolic array architecture. It is natural to arrange Kalman filter algorithms into a form of Faddeev algorithm to maximize the capability of hardware implementation of systolic arrays. First of all, this proposed algorithm avoids the direct matrix inverse computation as the usual back substitution in the Kalman Filter implementation, but obtains the values of desired results at the end of the forward course of computation, resulting in a considerable saving in added processing and storage. Secondly, since the Gaussian elimination procedure is applied through the computation, numerical stability is obtained. Thirdly, the parallel, modular computer architecture which consists of systolic processors provides simultaneous high throughput and a capability for a wide variety of linear algebraic operations.

In the following sections, we describe, in order, Kalman filters, the Faddeev algorithm, the implementation, a two-

dimensional systolic array architecture, and a numerical example.

## II. Kalman Filter

Kalman filters have been shown to be the optimal linear estimator in the least square sense for estimating dynamic system states in linear systems. The Kalman filter updates state estimation based on prior estimates and observed measurements. It consists of the model of the dynamic process which performs the function of prediction and a feedback correction scheme. The measurements can be processed as they occur, and there is no need to store any measurement data. However, all the associated matrices which describe the system dynamic, measurement system, and noise is assumed to be known. The conventional discrete time-varying Kalman filtering process involves the propagation of state estimates and error covariance matrices from time sample to next time sample. Discussions and applications on Kalman filters can be found widely in literature (Refs. 1 to 6).

The following equations define a general dynamic system and a measurement system.

$$\mathbf{x}(k+1) = \phi(k) \mathbf{x}(k) + \mathbf{w}(k) \quad (1)$$

$$\mathbf{z}(k) = \mathbf{H}(k) \mathbf{x}(k) + \mathbf{v}(k) \quad (2)$$

In Eq. (1),  $\phi(k)$  is a  $n \times n$  matrix called the state transition matrix which describes the plant,  $\mathbf{x}(k)$  is the state vector with  $n$ -dimension, and  $\mathbf{w}(k)$  is a  $n$ -vector called the disturbance. In Eq. (2),  $\mathbf{z}(k)$  is a  $m$ -vector termed the measurement vector;  $\mathbf{v}(k)$  is also a  $m$ -vector called the measurement noise vector, and  $\mathbf{H}(k)$  is a  $m \times n$  matrix called the measurement matrix. The disturbance  $\mathbf{w}(k)$  and noise  $\mathbf{v}(k)$  are assumed to be zero-mean Gaussian white noise sequences with covariance matrices  $\mathbf{Q}(k)$  and  $\mathbf{R}(k)$ , respectively. Furthermore, sequences  $\mathbf{w}(k)$  and  $\mathbf{v}(k)$  are assumed to be statistically independent. The Kalman filter is described by the following equations under assumptions that matrices  $\phi(k)$ ,  $\mathbf{H}(k)$ ,  $\mathbf{R}(k)$ , and  $\mathbf{Q}(k)$  are known.

$$\hat{\mathbf{x}}(k/k-1) = \phi(k-1) \hat{\mathbf{x}}(k-1/k-1) \quad (3)$$

$$\mathbf{P}(k/k-1) = \phi(k-1) \mathbf{P}(k-1/k-1) \phi^T(k-1) + \mathbf{Q}(k-1) \quad (4)$$

$$\mathbf{P}^{-1}(k/k) = \mathbf{P}^{-1}(k/k-1) + \mathbf{H}^T(k) \mathbf{R}^{-1}(k) \mathbf{H}(k) \quad (5)$$

$$\mathbf{K}(k) = \mathbf{P}(k/k) \mathbf{H}^T(k) \mathbf{R}^{-1}(k) \quad (6)$$

$$\Delta \mathbf{z}(k) = \mathbf{z}(k) - \mathbf{H}(k) \hat{\mathbf{x}}(k/k-1) \quad (7)$$

$$\hat{\mathbf{x}}(k/k) = \hat{\mathbf{x}}(k/k-1) + \mathbf{K}(k) \Delta \mathbf{z}(k) \quad (8)$$

$$k = 1, 2, \dots$$

Initial conditions are given as follows:

$$\mathbf{x}(0/0) = 0 \quad (9)$$

$$\mathbf{P}(0/0) = \mathbf{P}(0) \text{ or } \mathbf{P}^{-1}(0/0) = \mathbf{P}^{-1}(0) \quad (10)$$

Note that matrices  $\mathbf{P}(k/k)$  and  $\mathbf{P}(k+1/k)$  are commonly called error covariance matrices. The vector  $\hat{\mathbf{x}}(k/k)$  represents the optimal estimate of  $\mathbf{x}(k)$  based on the measurement sequence  $\{\mathbf{z}(1), \mathbf{z}(2), \mathbf{z}(3), \dots, \mathbf{z}(k)\}$ . Equations (3) and (4) are referred to as time updates and Eqs. (5), (6), and (8) are referred to as measurement updates. These filter equations are computed in order as listed from Eqs. (3) to (8).

## III. Faddeev Algorithm

Consider a linear equation:

$$\mathbf{AX} = \mathbf{B} \quad (11)$$

Suppose that it is desired to find  $\mathbf{CX} + \mathbf{D}$ , a linear combination of  $\mathbf{X}$  without first finding  $\mathbf{X}$ , which is an unknown. This can be represented in matrix form as:

$$\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline & \\ -\mathbf{C} & \mathbf{D} \end{array} \quad (12)$$

where  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$ , and  $\mathbf{D}$  are matrices or vectors. The dimension of the matrices will be discussed later. By adding suitable combinations of  $\mathbf{A}$  and  $\mathbf{B}$  to  $-\mathbf{C}$  and  $\mathbf{D}$ , the term " $-\mathbf{C} + \mathbf{WA}$ " appears in the lower left hand quadrant and " $\mathbf{D} + \mathbf{WB}$ " in the bottom right hand quadrant. It shows in matrix form as:

$$\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline & \\ -\mathbf{C} + \mathbf{WA} & \mathbf{D} + \mathbf{WB} \end{array} \quad (13)$$

If  $\mathbf{W}$  specifies the appropriate linear combination such that the lower left hand side of Eq. (13) is zero, then  $\mathbf{CX} + \mathbf{D}$  will appear on the lower right hand side.

That is,

$$W = CA^{-1} \quad (14)$$

$$\begin{aligned} D + WB &= D + CA^{-1}B \\ &= CX + D \end{aligned} \quad (15)$$

Representing the above equation in matrix form, we obtain:

$$\begin{array}{c|c} A & B \\ \hline 0 & CX + D \end{array} \quad (16)$$

Note that  $CX + D$  is the desired result and appears in the bottom right hand quadrant after the process to annul the bottom left hand quadrant.

It is clear from the above illustration that given matrices  $A$ ,  $B$ ,  $C$  and  $D$ ,  $CA^{-1}B + D$  can be computed. Consequently, by proper selection of matrices  $A$ ,  $B$ ,  $C$ , and  $D$ , the capability of the Faddeev algorithm can be fully utilized. The various possible matrix manipulations are shown in Fig. 1.

The simplicity of the algorithm is due to the absence of a necessity to actually identify the multipliers of the rows of  $A$  and the elements of  $B$ ; it is only necessary to "annul the last row." This can be done by triangularization, a numerically stable procedure, combined with an equally stable Gaussian elimination procedure (Refs. 12 and 13).

One of the more important features of this algorithm is that it avoids the usual back substitution or solution to the triangular linear system and obtains the values of the unknowns directly at the end of the forward course of the computation, resulting in a considerable savings in added processing and storage.

#### IV. Implementation

Kalman filter algorithms are adjusted to be the type of Faddeev algorithm in this section (Ref. 14). Computations are cyclically propagated through the following ordered set of passes. Note that the new data could be shifted into the array from the top, row by row as the calculation proceeds, so that there would be no delay in starting the next matrix computation.

1st pass:

$$\begin{array}{c|c} I & \hat{x}(k-1/k-1) \\ \hline -\phi(k-1) & 0 \end{array} \longrightarrow \hat{x}(k/k-1)$$

2nd pass:

$$\begin{array}{c|c} P^{-1}(k-1/k-1) & \phi^T(k-1) \\ \hline -\phi(k-1) & Q(k-1) \end{array} \longrightarrow P(k/k-1)$$

3rd pass:

$$\begin{array}{c|c} R(k) & I \\ \hline -H^T(k) & 0 \end{array} \longrightarrow H^T(k)R^{-1}(k)$$

4th pass:

$$\begin{array}{c|c} P(k/k-1) & I \\ \hline -I & 0 \end{array} \longrightarrow P^{-1}(k/k-1)$$

5th pass:

$$\begin{array}{c|c} I & H(k) \\ \hline -H^T(k)R^{-1}(k) & P^{-1}(k/k-1) \end{array} \longrightarrow P^{-1}(k/k)$$

6th pass:

$$\begin{array}{c|c} P^{-1}(k/k) & H^T(k)R^{-1}(k) \\ \hline -I & 0 \end{array} \longrightarrow K(k)$$

7th pass:

$$\begin{array}{c|c} I & \hat{x}(k/k-1) \\ \hline H(k) & z(k) \end{array} \longrightarrow \Delta z(k)$$

8th pass:

$$\begin{array}{c|c} I & \Delta z(k) \\ \hline -K(k) & \hat{x}(k/k-1) \end{array} \longrightarrow \hat{x}(k/k)$$

Note that the results (lower right quadrant) of each pass in general, needs to be stored and is used in later passes as new entries.

## V. Architecture

A general architecture is proposed to process all eight passes as mentioned in the previous section. To fit the type of Faddeev algorithm, the number of columns of the matrix (lower left quadrant) should be less than or equal to that of the matrix (upper left quadrant). Fortunately, all eight passes meet this requirement so that the matrix (lower left quadrant) will be annulled row by row. This can be done by ordinary Gaussian elimination. However, Gaussian elimination in general requires pivoting, and the pivoting strategy is not suited to a systolic array since it may require global communication for the pivot section. The neighbor pivoting technique introduces a zero to a row by subtracting a multiple of an adjacent row from it, interchanging the two rows when necessary to prevent the multiple from exceeding unity (Ref. 13).

This process is shown in the modified Faddeev algorithm (a two-step procedure):

$$\begin{array}{c|c} \text{A} & \text{B} \\ \hline -\text{C} & \text{D} \end{array} \xrightarrow{\text{step 1}} \begin{array}{c|c} \text{T} & \text{MB} \\ \hline -\text{C} & \text{D} \end{array}$$
  

$$\begin{array}{c|c} \text{T} & \text{MB} \\ \hline -\text{C} & \text{D} \end{array} \xrightarrow{\text{step 2}} \begin{array}{c|c} \text{T} & \text{MB} \\ \hline -\text{C} + \text{WT} & \text{D} + \text{WMB} \end{array}$$

where  $\mathbf{T}$  is an upper triangular, and  $\mathbf{M}$  is a nonsingular (triangularization) matrix. Since  $\mathbf{W} = \mathbf{CT}^{-1}$ , the final result is  $\mathbf{G} = \mathbf{D} + \mathbf{WMB} = \mathbf{D} + \mathbf{CA}^{-1}\mathbf{B}$ . In order to triangularize matrix  $\mathbf{A}$  and to annul matrix  $\mathbf{C}$ , it is necessary to have a two-step procedure. First,  $\mathbf{A}$  is triangularized by the neighbor pivoting process (simultaneously applied to  $\mathbf{B}$ ); second,  $\mathbf{C}$  is annulled by Gaussian elimination using the diagonal elements of  $\mathbf{T}$  as pivot elements. The square processor arrangement for both triangularization and annulling steps is shown in Fig. 2 (Refs. 10 and 11). However, to compute all eight passes, the size of the processor is  $2n$  cells (row) by  $2n$  cells (column).

It is desirable to have a fixed size processor to handle all eight passes. However, the size of the entry matrix/vector varies from pass to pass. By padding zeros in appropriate places, the  $2n$  cells (row) by  $2n$  cells (column) become the proper size for implementing Kalman filters. The scheme of zero padding is illustrated in Fig. 3. By using the fixed size  $2n \times 2n$  processor arrays, the estimate  $\mathbf{x}(k/k)$  can be updated in each  $16n$  time unit (assuming that it takes one time unit to operate data in a cell.)

## VI. Numerical Example

An example of air-traffic-control radar is chosen for numerical simulation. In this example, one has range, range rate,

bearing, and bearing rate as four components (in order) of the state vector  $\mathbf{x}(k)$ . The maneuver noise, which impacts on the change in range rate and bearing rate, is assumed to be zero-mean white, and the range rate and bearing rate are uncorrelated and with variances of 330 and  $1.3 \times 10^{-8}$ , respectively. The radar sensors are assumed to provide noisy measurements of the range and bearing. The noise of the radar sensors are assumed to be zero-mean white, and the range rate and bearing rate are uncorrelated and with variances of  $10^6$  and  $2.89 \times 10^{-4}$ , respectively. The values of the matrices and initial values are listed as follows:

$$\phi(k) = \begin{bmatrix} 1 & 15 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 15 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$H(k) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$Q(k) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 330 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.3 \times 10^{-8} \end{bmatrix}$$

$$R(k) = \begin{bmatrix} 10^6 & 0 \\ 0 & 2.89 \times 10^{-4} \end{bmatrix}$$

$$P^{-1}(0/0) = \begin{bmatrix} 0.2 \times 10^{-5} & -0.149 \times 10^{-4} & 0 & 0 \\ -0.149 \times 10^{-4} & 0.222 \times 10^{-3} & 0 & 0 \\ 0 & 0 & 0.662 \times 10^4 & -0.483 \times 10^5 \\ 0 & 0 & -0.483 \times 10^5 & 0.738 \times 10^6 \end{bmatrix}$$

$$\hat{x}(0/0) = \begin{bmatrix} 0.200 \times 10^4 \\ 0 \\ 0.500 \\ 0 \end{bmatrix}$$

This example was simulated on a VAX 780. The Kalman gain  $K_{11}(k)$ , the range error covariance  $P_{11}(k/k-1)$ , and the bearing error covariance  $P_{33}(k/k-1)$  were plotted against time index  $k$  and shown in Fig. 4. These plots show the numerical stability of the implementation of the Kalman filter on the concurrent systolic arrays.

## References

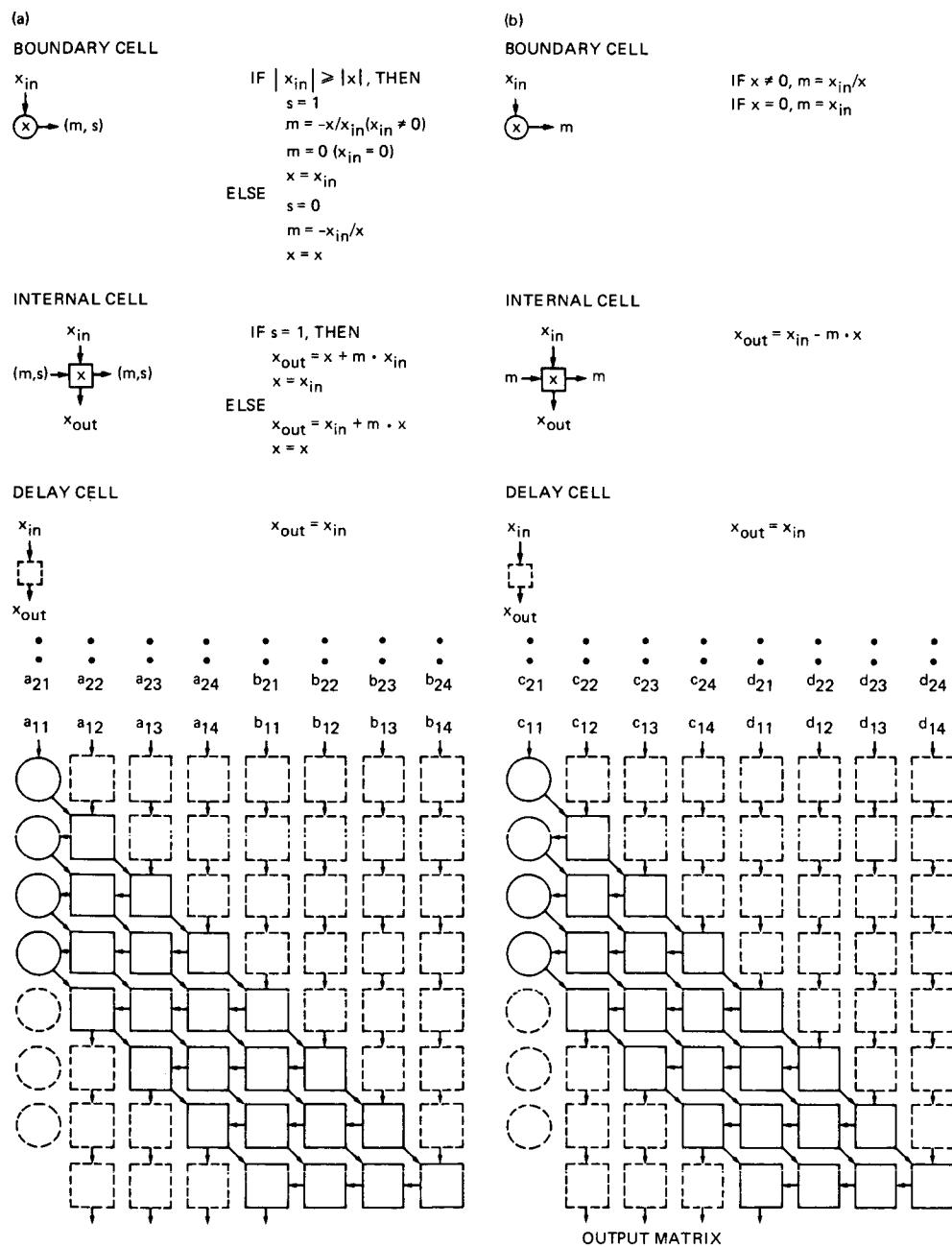
1. Gebb, A., *Applied Optimal Estimation*, The M.I.T. Press, Cambridge, Mass., 1974.
2. Meditch, J. S., *Stochastic Optimal Linear Estimation and Control*, McGraw-Hill, New York, 1969.
3. Yeh, H. G., "A Design Method for Failure-Proof Systems," *Proceedings of the 1983 American Control Conference*, San Francisco, Calif., June 1983, pp. 1219-1223.
4. Yeh, H. G., *Techniques for the Detection, Estimation, Distinction, and Compensation of Failures in Linear Systems*, Ph.D. Dissertation, U.C. Irvine, 1982.
5. Sorenson, H. W., *Parameter Estimation*, Marcel Dekker, New York, 1980.
6. Bryson, A. E., and Ho, Y. C., *Applied Optimal Control*, Rev. ed., Halsted Press, Div. of John Wiley & Sons, Waltham, Mass., 1975.
7. Yeh, C. S., Reed, I. S., Chang, J. J., and Truong, T. K., "VLSI Design of Number Theoretic Transforms for a Fast Convolution," *Proceedings of the 1983 IEEE International Conference on Computer Design: VLSI in Computers*, Port Chester, New York, Oct. 31 - Nov. 3, 1983, pp. 202-203.
8. Chang, J. J., Truong, T. K., Shao, H. M., Reed, I. S., and Hsu, I. S. "The VLSI Design of a Single Chip for the Multiplication of Integers Modulo a Fermat Number," *IEEE Trans. on ASSP*, Vol. ASSP-33, No. 6, Dec. 1985, pp. 1599-1602.
9. Kung, S. Y., Whitehouse, H. J., and Kailath, T., *VLSI and Modern Signal Processing*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1985, pp. 375-338.
10. Nash, J. G., and Hansen, S., "Modified Faddeev Algorithm for Matrix Manipulation," *Proceedings of the 1984 SPIE Conference*, San Diego, Calif., August 1984, pp. 39-46.
11. Nash, J. G., and Petrozolin, C., "VLSI Implementation of a Linear Systolic Array," *Proceedings of ICASSP, 1985*, Tampa, Florida, March 26-29, 1985, pp. 1392-1395.
12. Kung, H. T., Sproull, R., and Steele, G., *VLSI Systems and Computations*, Computer Science Press, Carnegie-Mellon University, Pa., Oct. 1981, pp. 367-378.
13. Gentleman, W. M., and Kung, H. T., "Matrix Triangularization by Systolic Arrays," *SPIE Real-Time Signal Processing IV*, San Diego, Calif., Aug. 1981, Vol. 298, pp. 19-26.
14. Yeh, H. G., "Kalman Filtering and Systolic Processors," *Proceedings of the IEEE International Conf. on Acoustic, Speech, and Signal Processing*. Tokyo, Japan, April 1986, pp. 2139-2142.

$$\frac{A}{-1} \left| \begin{array}{c|c} I & \\ \hline -C & 0 \end{array} \right. \longrightarrow A^{-1} \quad \frac{A}{-C} \left| \begin{array}{c|c} I & \\ \hline 0 & 0 \end{array} \right. \longrightarrow CA^{-1}$$

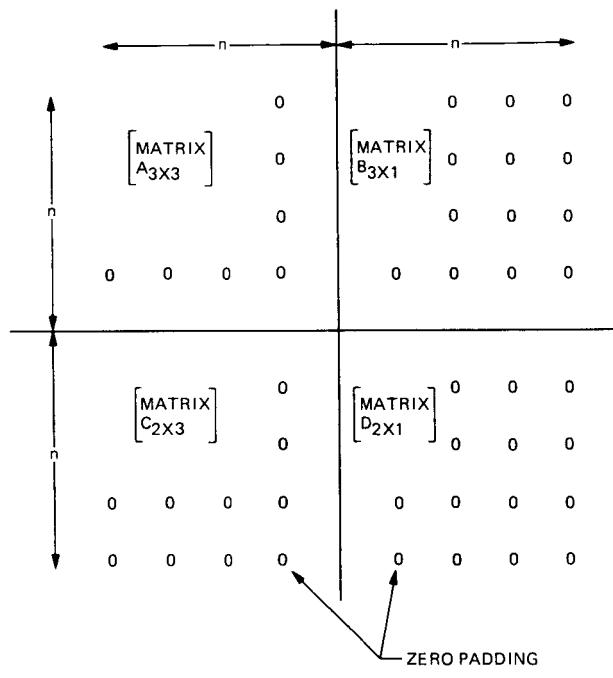
$$\begin{array}{c|c} I & B \\ -C & 0 \end{array} \longrightarrow CB \quad \begin{array}{c|c} A & B \\ -1 & 0 \end{array} \longrightarrow A^{-1}B$$

$$\frac{I}{-C} \left| \begin{array}{c} B \\ D \end{array} \right. \longrightarrow D + CB \quad \frac{A}{-C} \left| \begin{array}{c} B \\ D \end{array} \right. \longrightarrow CA^{-1}B + D$$

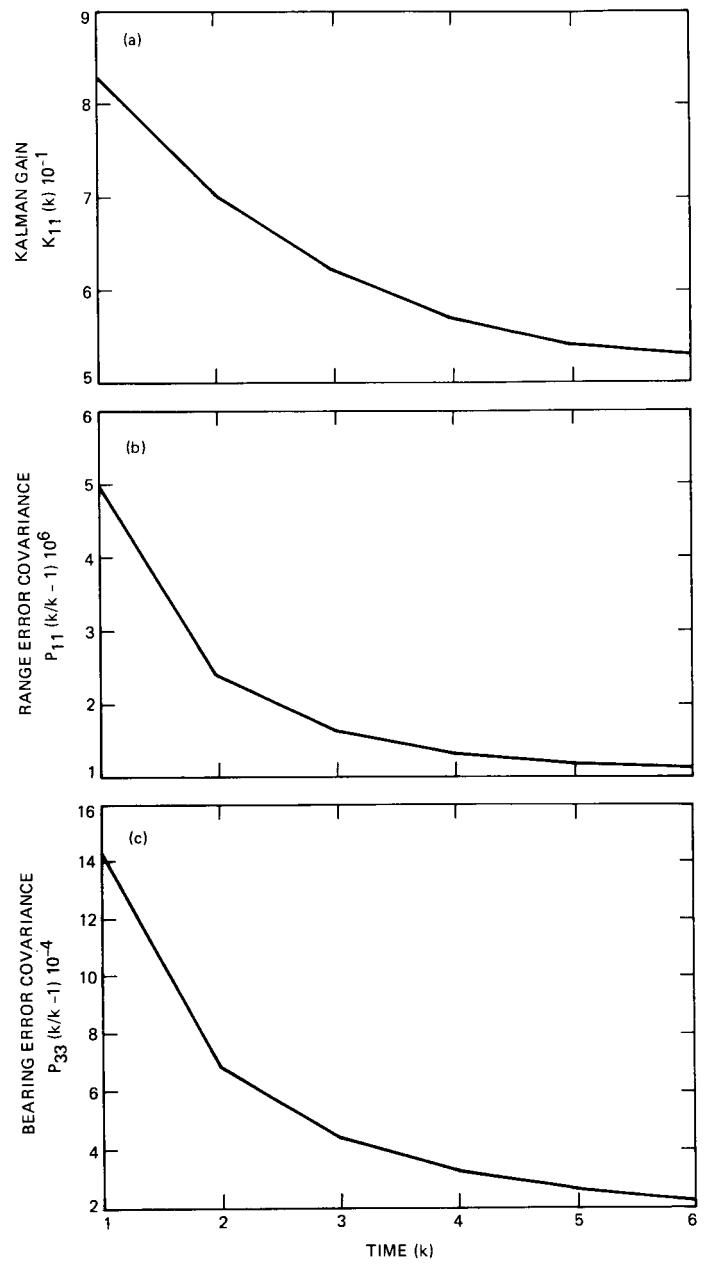
**Fig. 1. Examples of a variety of matrix operations possible with the Faddeev Algorithm**



**Fig. 2. Data Flow and PE arrangement:** (a) triangularization process,  $n = 4$ ; (b) annulling process,  $n = 4$



**Fig. 3. Zero padding scheme for implementing a small-size matrix/vector on processor arrays**



**Fig. 4. Computer simulation on the numerical computation of a Kalman filter: (a)  $K_{11}(k)$  vs time  $k$ ; (b)  $P_{11}(k/k-1)$  vs time  $k$ , and (c)  $P_{33}(k/k-1)$  vs time  $k$**